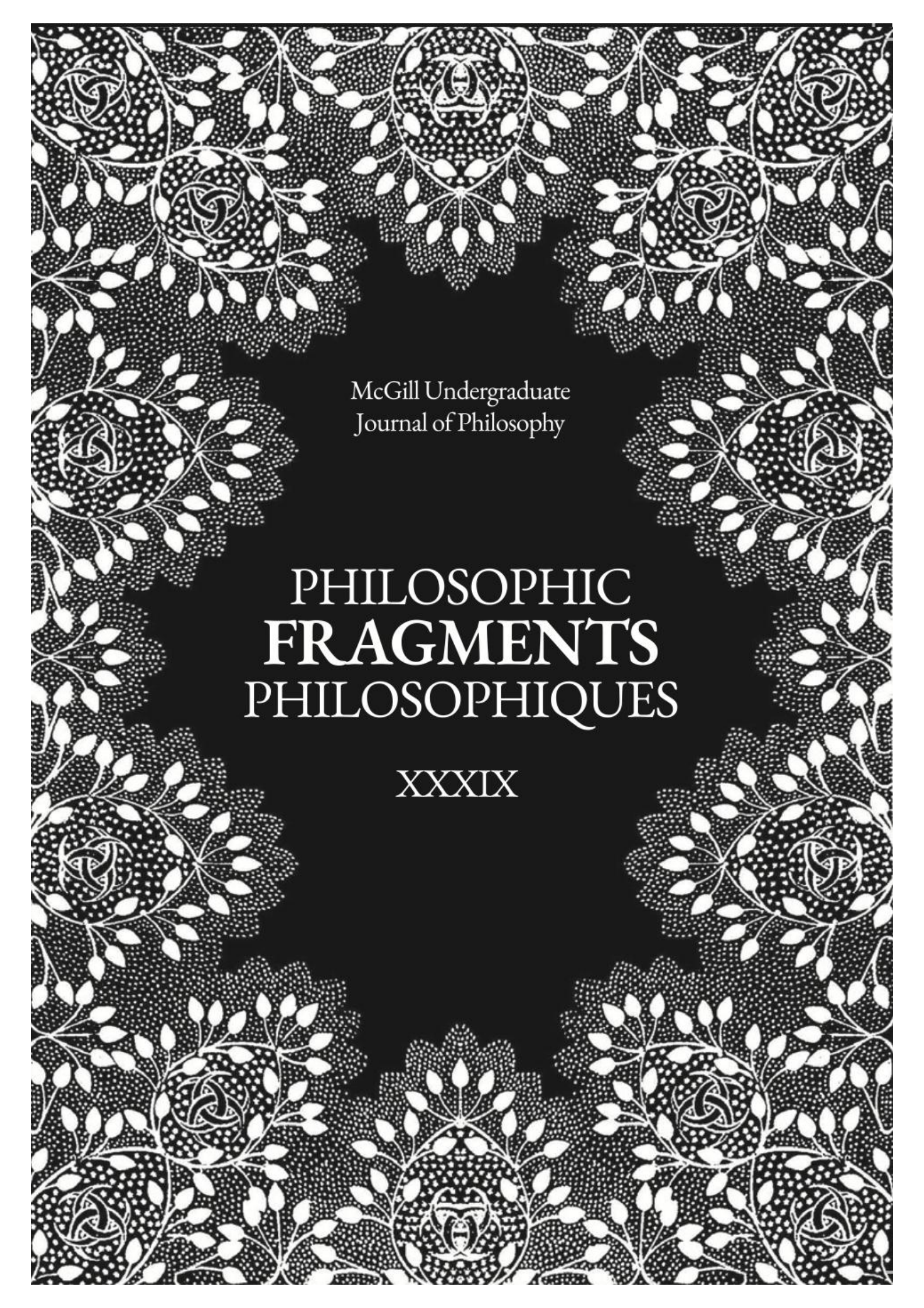




McGill Undergraduate
Journal of Philosophy

PHILOSOPHIC
FRAGMENTS
PHILOSOPHIQUES

XXXIX

PHILOSOPHIC
FRAGMENTS
PHILOSOPHIQUES

TEAM

EDITORS-IN-CHIEF:

Abby McCormick

Sofia Vaillant Forlini

Hadrien Velde

EDITORS:

Jenna Durante

Wenxi Geng

Camille Istamboulian

Hazel Krauth

Gregor McCall

COVER DESIGN & PHOTOGRAPHY:

Rohan Khanna

*This project was generously funded by the AUS Financial
Management Committee through its Journal Fund.*

CONTENTS

THE CASE FOR AN INTERSECTIONAL APPROACH TO AFFIRMATIVE ACTION: COMPREHENSIVE JUSTICE

MARK XIAO 4

WHAT WE TALK ABOUT WHEN WE TALK ABOUT ARTS: A LINGUISTIC INQUIRY ON AESTHETICS

RUTH WANG 14

LONGTERMISM AND ARTIFICIAL INTELLIGENCE: AN ANALYSIS OF FUTURE WORLDS

SABRINA SAVOIA 28

THE INHERENT STRANGENESS OF MORAL COHERENCE

LINDSAY COBB 40

VEILING ILLUSIONS: LANGUAGE AND REASON AS VEILS CONCEALING THE TRUE ESSENCE OF THE WORLD

ANNA BISTOUR 52

ON THE POSSIBILITY OF NONVIOLENT DECOLONIZATION: FANON, IN LIGHT OF GHANDI

VALENTIN PELOUZET 62

AMOR MUNDI: POLITICAL METAMORPHOSIS OF LOVE

JULIAN MINICONI 82

FANON IN THE CONSULTING ROOM: COLONIALISM, RACISM, AND THE URGENT NEED FOR A MORE INCLUSIVE MENTAL HEALTH-CARE SYSTEM

ANAHAT GHUMAN 96



The Case for an Intersectional Approach to Affirmative Action: Comprehensive Justice

MARK XIAO

This paper will argue against Race-Based Affirmative Action (RBAA) in college admissions on the basis that RBAA fails to charitably represent applicants, which is unjust. As an alternative, this paper will propose an ‘Intersectional Affirmative Action’ (IAA) based on Crenshaw’s 1991 work on intersectionality. IAA is uniquely comprehensive: it can serve as a framework for comparing and understanding applicants as multifaceted individuals. IAA considers both applicants’ collective and individual identities, while Traditional Affirmative Action selectively omits key parts of an applicant’s identity. For example, a Black female applicant would face an unjust and reductive characterization under RBAA, which sees her only as a Black person, or under gender-based Affirmative Action, which sees her only as a woman. IAA can fully acknowledge her identity as a Black woman and help her to be understood as such. Given the precedent for systems, such as the U.S. sentencing guidelines, to quantify characteristics such as race and gender into scores that factor into life-changing decisions, IAA can be practically implemented. It is possible to create a similar system, guided by intersectionality, that produces a score that can act as another factor in college admissions, such as SAT scores and GPA calculations.

Keywords: race, justice, intersectionality, Affirmative Action.

Affirmative action refers to a set of policies designed to improve “educational opportunities for members of minority groups” by giving “limited preferences to minorities... in admission to institutions of higher education,” as defined by Encyclopedia Britannica. These policies aim to right historical injustices, which helps to increase the admitted class’s diversity in a given context. In university admissions, Affirmative Action has fixed race as the main criterion for evaluating minority applicants. I shall refer to this kind of Affirmative Action as ‘Race-Based Affirmative Action’ (RBAA). RBAA in university admissions should not be conflated with Affirmative Action in the workplace or job market, as each addresses different issues and has different aims. I will evaluate the impacts of Affirmative Action exclusively as it has appeared in university undergraduate admissions.

Recently, the U.S. Supreme Court ruled that RBAA in some U.S. universities is unconstitutional because it “hurt some races and helped others” (Waxman). The debate for and against Affirmative Action persists, with some clamouring for the return of RBAA or similar policies. In this paper, I will argue against Race-Based Affirmative Action because it is too narrow to accomplish its purpose of addressing discrimination and righting historical wrongs. It also encourages the omission of relevant facts about the applicants, perpetuating injustice. As an alternative, I will advocate for implementing Intersectional Affirmative Action (IAA), based on Kimberlé Crenshaw’s writings on intersectionality, because it offers a clear solution for comprehensive applicant representation. I will then analyze a counterargument involving the practicality of IAA but respond that there already is a precedent for systems that can translate social inequalities into standardized scores that may then be used to consider each individual’s case.

My claim that RBAA constitutes a system of injustice may immediately seem to some to be baseless and novel. To elucidate my position, I will first introduce a conception of ‘Justice’ and how it universally functions both in university admissions and court decisions. A comprehensive definition of ‘Justice’ is beyond the scope of this paper, so for the sake of brevity, I will be utilizing the common definition of ‘Justice’ (as can be found in any dictionary), which essentially amounts to ‘fairness.’ I would like to emphasize its foundation in equity—the core principle of Affirmative Action, which seeks to redress past and present discrimination through equitable support.

Nowhere is the concept of ‘Justice’ more important than in the law and

the courts of law that uphold it. In legal proceedings, justice has a fundamental prerequisite: truth. This notion is reflected in the emphasis on facticity—the pursuit of objective truth—underpinning judicial decisions. For example, in a Canadian court of law, when a witness is sworn in for their testimony, they are always compelled to swear, “I solemnly affirm that the evidence to be given by me shall be the truth, the whole truth and nothing but the truth” (Canadian Evidence Act). In an American court of law, this oath is largely unchanged except that the witness may invoke a religion of their choice. Either way, it is clear that, in North America, the concept of an absolute, unobscured truth is a fundamental part of ‘Justice.’

Logically speaking, a court system that fails to consider all pertinent evidence before reaching a verdict could never be said to be ‘fair’ because that decision is predicated on incomplete evidence. An informed, just decision cannot be made when factuality is even partially obscured. The concept of ‘lying by omission’ demonstrates the possible issues with obscuring relevant information, as it can be just as harmful and counterproductive toward reaching a just or fair decision as lying outright. Hypothetically, if a man was tried for a serious crime, and that court did not even consider evidence that may have exonerated him, that would be unjust because they did not do their due diligence (which I define specifically as the level of diligence needed to represent facts of a person’s situation most accurately) in proving the guilt of the defendant. In a court of law, due diligence is expected, lest an injustice be committed.

RBAA is similarly unjust to this court because it leads the decision-makers to fail to consider all the available evidence that may sway a life-changing decision. RBAA fails to comprehensively evaluate all of the relevant ways that an applicant may have faced barriers to their education. For example, RBAA does not consider factors such as gender, sexual orientation, class, or economic circumstances. Thus, RBAA fails to achieve justice because, by definition, it does not ensure well-informed decisions. Since true justice is an enclitic of a whole truth, RBAA is wrong in the way that lying by omission is wrong and counterproductive towards justice in the same way.

As a rule of thumb, admissions committees must understand the struggles an applicant may face and how they affect the applicant’s application profile. Suppose we create Affirmative Action policies that could sway the outcome of whether a given applicant can attend a university. In that case, it is arguably better, both for the interests of those being evaluated and those evaluating, to know

more about the applicant. The best decisions are the most well-informed; ignorance is the enemy of true ‘Justice.’ In the end, I argue that any Affirmative Action policies that only focus on one aspect of an applicant’s identity never present a comprehensive understanding of their background or any struggles they may have faced as minorities.

Moreover, by viewing an applicant’s life exclusively through a race-centric lens, we simultaneously omit and discount any other struggles they may have faced from their different identities, such as those pertaining to their gender or economic status. While gender- and economic-based Affirmative Action movements also exist, the pursuit of ‘Justice’ requires an approach that comprehensively acknowledges an individual’s lived experience. Thus, it makes more sense to integrate these considerations under IAA, as I will discuss later. Additionally, if we truly seek fairness, we ought to take an intersectional approach because none of these perspectives are mutually exclusive.

For example, as Crenshaw notes, a Black female applicant would face discrimination not only as a Black person but as a woman and, additionally, as a Black woman (Crenshaw, *Mapping the Margins*, 1244). She faces a compound kind of oppression that cannot be reduced to that of a single category or identity. As such, it is neither ‘Fair’ nor ‘Just’ for a policy meant to right historical wrongs to only judge this applicant on the basis of her skin colour when she has faced barriers (related to these historical wrongs) due to other aspects of her identity. In the interest of fairness, all of her struggles, within reason, should be considered, as to consider some, but not all, of her struggles is simply disingenuous and represents a lack of due diligence.

A commonly proposed alternative to RBAA, Class/Economic Based Affirmative Action, similarly falls prey to this pernicious problem of underrepresenting the barriers applicants face. Moreover, both fail to account for how these barriers ‘intersect’—how the identities of being Black and female can compound and cause a unique form of oppression. Just as the justice system requires court officers to do their due diligence in their duties, the onus falls upon the admissions committee to accurately and fairly evaluate their assigned applicants—and this cannot be done if they ignore some or all of the latter’s struggles. Now, we stand at a metaphorical crossroads. We can continue forward with the current abolition of all Affirmative Action, recreate RBAA or any other single-criteria Affirmative Action policy, or choose a new option to replace the abolished RBAA—an op-

tion I will advocate for. I will not create another novel argument for Affirmative Action: all that I will argue is that, as previously established, 'Justice' is never achieved by shirking or altogether ignoring the truth, whether intentionally or not. Thus, abandoning all forms of Affirmative Action is counterproductive to being 'Just.' Of course, even a half-hearted attempt at achieving equity is better than no attempt, but 'Justice' and equity should never be half-hearted.

This leads to the fundamental question: What would be a more 'Just' system that could replace RBAA or any kind of single-axis Affirmative Action? I advocate for a more holistic and comprehensive approach towards Affirmative Action. Minorities exist in all kinds of demographics, by definition, whether they pertain to race, gender, sex, sexual orientation, or any other such category. Many of these minorities have suffered and continue to suffer identity-based discrimination and, in many cases, find that these discrete, identity-based forms of discrimination intersect. With all this in mind, I present an intersectional form of Affirmative Action (IAA) as the best option based on Crenshaw's definition of intersectionality. Crenshaw writes, "Discrimination, like traffic through an intersection, may flow in one direction... an accident... can be caused by cars travelling from any number of directions and, sometimes, from all of them. Similarly, if a Black woman is harmed because she is in the intersection, her injury could result from sex discrimination or race discrimination" (Crenshaw, *Demarginalizing the Intersection of Race and Sex*, 149). Essentially, she likens the cars to different, discrete forms of discrimination and shows that these forms can 'intersect,' or in other words, pile upon a person like a pile of cars in a traffic accident. Intersectionality is especially relevant to Affirmative Action, as Crenshaw argues that "the problem with identity politics...[is] that ... it frequently conflates or ignores intra-group differences" (Crenshaw, *Mapping the Margins*, 1242). This is to say that even within broad and arbitrary groups, like race or gender, there is great variation in the lived experiences of these people. A truly just system must reflect this complexity instead of reducing these people to singular aspects of their identities.

For example, compare the cases of a Chinese national male student from the heart of metropolitan Beijing and a Chinese-Canadian male student who spent his entire life living in Vancouver's Chinatown. The latter area is known as one of the "poorest areas of Vancouver" (Yamamoto et al.). The former is privileged to be able to easily afford the expensive international tuition, which is juxtaposed with the latter, who has had more barriers towards obtaining his education and life

in general and would be barely able to afford the domestic rate for tuition, much less the international tuition rate.

An intersectional framework can offer a focal point for evaluating all of these different factors. With intersectionality in mind, the admissions committee could look deeper into how the latter applicant's financial status may have intersected with his racial identity and how that may have impacted his life. Similarly to the case of the Black female applicant, IAA serves as an impetus to scrutinize how race- and gender-based discrimination can intersect. After all, as Crenshaw notes, "racism and sexism readily intersect in the lives of real people" (Crenshaw, *Mapping the Margins*, 1242). Compared to the other options available through the aforementioned 'crossroads,' it is clear that IAA would be the most comprehensive and fair way to evaluate applicants' struggles and to right historical wrongs. Now, given that I am advocating for a policy change, I will do my due diligence and anticipate the most convincing possible argument against my Intersectional Affirmative Action (IAA). This argument falls under concerns of practicality—namely, it is that injustice is arguably difficult to measure, which means that IAA would greatly complicate the admissions system by increasing what the admissions committees must consider, perhaps to the extent of placing an undue or unrealistic burden upon said committees.

My answer is that a precedent to quantify theoretically 'difficult to measure' social circumstances already exists in the United States Sentencing Guidelines, which assign numerical 'scores' to offenders that can be adjusted based on various factors such as their criminal history, history of addiction, mental state, and age (United States Sentencing Commission). These scores affect the severity of sentencing. As such, it cannot be said that it is impossible to measure injustice. With a few modifications, this system could be adapted to evaluate college applicants and assign them a standardized score by considering different factors—much like how actuaries calculate insurance premiums, which notably include the consideration of gender (Financial Consumer Agency of Canada). Then, admissions committees could produce a standardized score that will act as a single factor in the sea of factors an admissions committee must examine to make admissions decisions. Even if this factor is not intended to be the sole determinant of an applicant's success, at the very least, IAA allows for a more comprehensive and just representation of a given applicant. Although I predict that this system will likely have some flaws as it develops, it arguably facilitates the most com-

prehensive possible form of ‘Justice’ —one that is sympathetic to the existence of different factors in an applicant’s identity and how they intersect.

‘Justice’ demands precision and not shortcuts—the ideal of due diligence is key. In the interest of being ‘fair,’ as established earlier, the onus falls squarely upon the shoulders of the admissions committees to ensure that ‘Justice’ is carried out (within reason). RBAA and similarly single-minded strains of Affirmative Action fall pitifully short of the intention to charitably and accurately represent an applicant’s struggles. Instead, the best option is to utilize Crenshaw’s intersectionality approach and apply it to Affirmative Action to create a new system altogether. Just as paradigms shift in science due to the ever-changing needs of humankind, so too must paradigms shift in public policy as social needs ought to be better addressed. The first step is to recognize that ‘Justice’ is mutually exclusive with ignorance.

Works Cited

- Crenshaw, Kimberlé. “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.” *University of Chicago Legal Forum*, vol. 1989, no. 1, Jan. 1989, p. 139.
- Crenshaw, Kimberlé. “Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color.” *Stanford Law Review*, vol. 43, no. 6, July 1991, pp. 1241–99, <https://doi.org/10.2307/1229039>
- Department of Justice Canada. “Canada Evidence Act.” *Government of Canada*, 1985, laws-lois.justice.gc.ca/eng/acts/c-5/page-1.html
- Financial Consumer Agency of Canada. “Determining Your Insurance Needs.” *Government of Canada*, 22 Aug. 2023, www.canada.ca/en/financial-consumeragency/services/insurance/determining-insurance-needs.html
- The Editors of Encyclopedia Britannica. “Affirmative Action: Definition, History, & Debate.” *Encyclopædia Britannica*, 12 Feb. 2025, www.britannica.com/topic/affirmative-action
- United States Sentencing Commission. “Sentencing Guidelines”. *United States Sentencing Commission* www.ussc.gov/sites/default/files/pdf/about/overview/2022_Guidelines-BasicsTrifold.pdf
- Waxman, Olivia. “Supreme Court Rules against Race-Based Affirmative Action.” *Time*, 29 June 2023, time.com/6291182/affirmative-action-supreme-court-decision-overturs/
- Yamamoto, Sho, et al.. “Spill the Tea: Gentrification of Vancouver Chinatown.” www.sfu.ca/communication/news-and-community/blog/spill-the-tea-gentrification-of-vancouver-chinatown.html



What We Talk About When We Talk About Arts: A Linguistic Inquiry on Aesthetics

To Caryn Azoff—An Artist and Teacher

RUTH WANG

What do we mean when we call something ‘art’ or ‘beautiful’? In a time when people rarely agree on these terms—if they ever did—it is worth asking whether our confusion comes not from the things we are trying to describe but from the words we use to describe them. This paper explores how language in aesthetic appreciation shapes our understanding of art and beauty. Drawing on Ludwig Wittgenstein’s later philosophy, especially his reflections on how language functions in everyday life, I argue that aesthetic appreciation can only be interpreted within its linguistic context. Using examples from music, painting, and cross-cultural appreciation, I show that language not only interprets art but sometimes transforms or even fabricates its meaning. Ultimately, I question whether we even need to use aesthetic terms like ‘art’ or ‘beauty’ as much as we do and suggest that paying more attention to our ordinary experiences—and less to abstract labels—might bring us closer to what really matters in life.

Key words: philosophy of language, Ludwig Wittgenstein, aesthetics, art appreciation, art history, value judgment, contextualism, cultural context, public media.

What is ‘Art’? And what is ‘Beauty’? These questions seem particularly difficult in the contemporary age. We no longer agree on what we call ‘beauty,’ if indeed we ever did, nor can we draw even an ambiguous boundary of ‘art.’¹ Philosophers and laypeople alike severely question, if not shatter, the seemingly obvious correspondence between the two—namely, that art expresses or evokes a sense of beauty. While such problems seem to have only emerged with modernity, they reveal issues that have always



The Ugly Duchess, 1513. Quentin Matsys. Oil on oak panel. National Gallery, London.

existed within our particularly opaque aesthetic concepts. One reason behind our confusion amounts to our use of aesthetic vocabulary such as ‘beauty’ and ‘art.’ Another is that the persuasive narratives used to describe and interpret beauty and art—information plaques in galleries, art critiques, and art history—often change our way of seeing the same piece and sometimes completely overturn our attitudes toward them. Words surround art and introduce it to the viewer. They endow artworks with more meaning than what is directly in front of us, yet we often overlook their power and mistake the charm of language for the mystery of art.

In this paper, I will briefly examine the application of the terms ‘beauty’ and ‘art’ and discuss how people use language differently when talking about art. I wish to draw attention to the enchantment of aesthetic language and to question whether the chaos surrounding aesthetic concepts lies in the language we use around them. I take the approach of Ludwig Wittgenstein’s investigation of language, discuss his view on aesthetics, and move beyond it to criticize common but potentially misleading narratives surrounding art.

¹ Readers can find relevant discussions in the collection of essays, *Contemporary Debates in Aesthetics and the Philosophy of Art*, edited by Matthew Kieran.

To talk about aesthetics and aesthetic language, we should first clarify what exactly we are studying and how we should study it. Wittgenstein argues that all philosophical problems should be “solved through an insight into the workings of our language” (*Philosophical Investigation*, 109). Thus, looking into his discussions on aesthetics alongside his philosophy of language shows us a way to understand the subject.

Though Wittgenstein left few philosophical writings about aesthetics, this does not imply that he had no personal interest in the subject. On the contrary, art, especially music, was important in his life (Glock, 31). Because Wittgenstein’s early views on language rejected the idea that propositions could be made about things “outside the world,” he had reasons to avoid putting aesthetics at the centre of his philosophical interests (*Tractatus Logico-Philosophicus*, 6.41). In the *Tractatus Logico-Philosophicus*, he briefly touches on aesthetics, suggesting that it is transcendental and “cannot be put into words” (6.421). Yet, in Wittgenstein’s later discussions, he mentions art and observes that “the subject of aesthetics is very big and entirely misunderstood” (*Lectures and Conversations*, 1). He argues that the frequent use of adjectives, such as ‘beautiful,’ ‘fine,’ and ‘lovely,’ in the context of aesthetic appreciation might mislead us into perceiving that aesthetic value is some inherent quality of things. We should be careful not to presume from the structure of a sentence such as “This is beautiful,”—which is “generally just subject and predicate”—that these words bestow meaningful qualities upon the ‘things’ to which they are applied (Wittgenstein, *Lectures and Conversations*, 5). Rather, we should redirect our attention to how this vocabulary plays a role in concrete life situations. As Wittgenstein argues in the *Philosophical Investigations*, to talk about what a word means, one should look at how it is applied and the function it serves in life activities (1–88).

Under such investigation, Wittgenstein finds that aesthetic adjectives are usually used as interjections or expressions of approval “as a substitute for a facial expression or a gesture” (*Lectures and Conversations*, 5). What makes them interjections of approval depends on the occasion in which they are said. Aesthetic phrases merely function to direct the audience’s attention to a particular work of art or natural beauty. However, in knowing the function of aesthetic expressions, the expressions themselves have “almost a negligible place” compared to the specific situations in which they are said (5). Unlike extra-aesthetic language, which usually plays “a characteristic part of a large group of activities,” such as

commanding another to do something, aesthetic judgments are “entirely uncharacteristic” (5). As interjections, plain praise — “This is beautiful!” — is no more effective than a pleasant facial expression and exclaiming, “Wow!” The latter may even be more versatile in terms of subtlety in expression and tone of voice.

Yet, most of our aesthetic appreciation consists not simply in approval or disapproval, but of characterizing a work of art (e.g., “This melody is melancholic”), noticing a specific detail (e.g., “Listen to the cello part”), making comparisons (e.g., “The beginning and the end use the same melody, yet have different moods”), and suggesting adjustments (e.g., “This part should be played softly”). Making aesthetic critiques of an object is usually more than expressing mere dislike but rather expressing what one desires it to be, whether out of personal preference or in conformity to social standards. For example, “This suit is too short,” simply expresses the commentator’s wish to have it be longer. Even if someone wants to show their approval of something, they often show it better in gestures and behaviours. Take Wittgenstein’s example: “How do I show my approval of a suit? Chiefly by wearing it often, liking it when it is seen, etc.” (*Lectures and Conversations*, 13). In other cases, appreciation can be “shown not by interjections, but by the way [one] chooses, selects, etc.” (19). Aesthetic appreciation takes place in a variety of forms. Appreciating a typical work representative of a certain style is “entirely different” from talking about those “tremendous” masterpieces that may not comply with the standards of their time yet still move us (23). In these various types of activities associated with aesthetic appreciation, aesthetic adjectives are hardly necessary or “play any role at all” (8).

Furthermore, focusing on the function of aesthetic expressions sheds light on the “enormously complicated situation[s]” in which they are used (Wittgenstein, *Lectures and Conversations*, 5). We often cast aesthetic judgments under very different social and cultural backgrounds, and in such cases, the context matters more in determining what these judgments mean than the specific words used. “To describe what [appreciation] consists in, we would have to describe the whole environment” (20). An appreciator from England might claim that he appreciates African Art but may perceive it in an “entirely different” manner compared to a person from an African cultural background (28). Without being immersed in the relevant culture and sharing its community’s form of life, one might not be aware of its artistic traditions. In this sense, one cannot judge a certain piece of work from an external perspective without knowing the aesthetic standards of its cul-

ture.

I will now further discuss the cultural contexts of art as they have greater social implications beyond individual art appreciation. Historically, appreciating without understanding a piece's cultural context has led to exoticism and cultural exploitation. Frédéric Chopin intended to show his Polish nationalism in his *polonaises* and *mazurkas*, yet he could not stop his audience in Parisian salons from regarding them as exotic pieces. Antonín Dvořák was inspired by Czech and East European folk songs, yet his works were interpreted as showing Slavic 'flavour' by his European and American listeners. Were *chinoiseries* appreciated properly without their audiences knowing the cultural significance of bamboo as a symbol of virtue and integrity? Or was placing Chinese porcelains in one's drawing room a mere display of the owner's wealth and taste? What are we viewing when we look at the transformation masks of Native Americans, stripped away from their rituals and now put in an art gallery? Without taking care of our language, all the differences in 'appreciation' are covered up by the object's 'aesthetic value,' as if it were apolitical, acultural, and ahistorical. What value are we talking about? Without careful attention to language, the rich historical and cultural significance of an object is reduced to mere 'aesthetic value,' as if it alone were a sufficient explanation for its worth. But what does 'aesthetic value' truly mean? Unlike historical or cultural values, which are grounded in an object's role, function, or the good it has done, aesthetic value is an illusion: it suggests a 'value' beyond its utility, as if something is still 'aesthetic' even if it does not give people pleasure, does not record history, does not express a certain political stand, does not reflect human conditions, does not remind people of their past experiences.

When making aesthetic judgments, our 'taste' depends on our life experience, our culture, history, and the political situation we are in. When we say something has aesthetic value, it is because it has served a specific function in our lives. Art does not exist by itself, away from people's lives. When people disagree on beauty or art, they are disagreeing on their experiences and memories. When someone



Fig.2. Tondal's Vision, mid-16th century. Follower of Hieronymus Bosch. Oil on panel. Lázaro Galdiano Museum.

sees beauty in something that another doesn't, it is because the former has a specific experience associated with it or because the object served some other role in their lives, such as entertainment or comfort. When we elevate 'aesthetic value' to the same importance as tangible values—such as utility, comfort, or ethical impact—we risk allowing the idea of aesthetics to overshadow historical, social, and ethical considerations. Yet, without taking care of our language, the diverse reasons behind 'appreciation' get covered up under the label of an object's 'aesthetic value,' as if it were apolitical and ahistorical. But what value are we really talking about? Do we accept something as having 'aesthetic value' just because enough people have declared it 'beautiful'?

Nevertheless, we can legitimately ask: is it impossible to 'appreciate' works of art from other cultures and periods? To what extent can one learn the aesthetic standards of another culture? Even after studying a local group's instruments and immersing themselves in its culture, an ethnomusicologist may still not internalize its musical traditions and cultural associations. A sympathetic response might be that, in a certain sense, one can still appreciate works of art separate from their contexts. According to the Death of the Author theory, meaning is created through the interaction between the work and the audience, beyond the limitation of the author's intentions and biographical background. Without such constraints, this allows for a multiplicity of interpretations (Barthes, 147). Though the audience's interpretation might deviate from the artist's original intentions, it can still produce inspiring ideas. Still, one should at least be aware of the cultural context of whatever one is 'appreciating' and keep in mind that they will necessarily interpret the work differently from both the artist themselves and the people of the artist's culture. A modern viewer might be naturally inclined to interpret a 16th-century painting of Hell symbolically and find it hard to ignore its similarities to 20th-century Surrealism despite knowing that, for the believers at the time, Hell was a physical place; the sight of the devil, fire, and human tortures would have evoked their most realistic fears. A Renaissance portrait of the Virgin Mary may have once been part of religious worship when it was first created, yet know-



Fig. 3. Virgin of the Rocks (Paris), 1483–1486. Leonardo da Vinci. Oil on panel (transferred to canvas). Louvre, Paris.

ing this does not prevent people in a secular society from viewing her through the lens of irreligious humanism. Our direct experience of an image is immediate. The moment we see a piece of art, we already form an impression of it—without thinking, processing, or analyzing it. Even if we understand the cultural context, we still have this raw, instinctive reaction when we first encounter the artwork, that is, how we personally feel at that moment. Our personal experiences of an artwork, though shaped by our knowledge and beliefs, cannot be replaced by the knowledge of its cultural context or historical records. Our first-person experience still constitutes what we call ‘aesthetic activity,’ even though it is not the only or always the most central part when we view art.

Confusion about aesthetics not only arises from abstracting it beyond its contextual use as a linguistic expression; words used to interpret the meaning of an artwork also often add more mystery than clarity. Even within the same cultural context, one might be surprised at how our attitudes can be affected by the text around images. Take John Berger’s example in his *Ways of Seeing*: “This is a landscape of a cornfield with birds flying out of it. Look at it for a moment. Then turn the page” (Berger, 27).



Fig. 4a. Wheatfield with Crows, July 1890. Vincent van Gogh. Oil on canvas. Van Gogh Museum, Amsterdam.



This is the last picture that Van Gogh painted before he killed himself.

Fig 4b.

Undoubtedly, the words have changed the meaning of the painting. “The image now illustrates the sentence” (28). One can find similar cases in music. Listen to Felix Mendelssohn’s *Overture Op. 2*. Listen closely to the soft violin at the beginning of the piece. Now, if the title of the piece, “A Midsummer Night’s Dream,” were revealed to the listener, it would seem obvious that the violins imitate the movement of Shakespeare’s flying fairies. But was it ‘obvious’ the first time one listened to it, without knowing beforehand the story intended for the piece? If the above examples are benign, there are also cases where words are intentionally chosen to disguise the purpose of an image.

Look at the nude in Figure 5, with the title *Love Seduces Innocence, Pleasure Entraps, and Remorse Follows*. From the title alone, one can almost convince oneself that the painting is meant to teach women about chastity while overlooking its visually explicit, opposite theme—nudity and eroticism for the sexual enjoyment of its male viewers. Do these words, juxtaposed with artworks, reveal the truth to us? Or are they also factors in our glorification



Fig. 5. Virgin of the Rocks (Paris), 1483–1486. Leonardo da Vinci. Oil on panel (transferred to canvas). Louvre, Paris.

of suffering and romanticization of the exotic other?

Similarly, reading an art critic's studies on paintings and music might alter our feelings about a piece of artwork or art as a whole. The ambiguity inherent in aesthetics then becomes a good excuse for mystifying art. We often associate art with vocabulary like beauty, truth, genius, civilization, status, taste, etc. (Berger, 11). But, many of these assumptions may intimidate people and discourage them from engaging with art in their daily lives. They make what is considered 'High Art' seem exclusive or difficult to access. Someone may feel they cannot 'properly appreciate' these forms of art because they are seen as requiring special knowledge or status. They distance people from various artistic activities—as if they are not qualified to understand the arts—while in reality, human beings, regardless of their social status, have always participated in and created what we now call 'art,' just as naturally as any other life activity. People invented 'rituals' for galleries, museums, theatres, and concert halls—conventions that set how one ought to behave on these occasions—and forgot that these 'rituals' are merely community customs and do not imbue the activities happening in these places with intrinsic value or essence. When Marcel Duchamp signed a urinal and claimed this 'readymade' as 'art,' he presumably intended to annihilate the boundary between art and everyday life. In a sense, he was right—if we examine our aesthetic language, no such boundary can be drawn in the first place. Yet he argued, "An ordinary object [could be] elevated to the dignity of a work of art by the mere choice of an artist" (The Museum of Modern Art). If there is such a thing as what he called the "dignity of a work of art," then this 'dignity' must be something so fragile that its very existence relies on the space of art exhibitions and the words of authorities (The Museum of Modern Art). Art puts itself in an awkward situation exactly by stating its unique value in ordinary life. The more it dignifies and isolates itself, the more it requires to be worshiped in galleries rather than appreciated in communal forms of life, to survive under the name of 'art.' If art is merely that which artists and critics call art, it becomes a game of wealth and power. After all, outside of social and political structures, under what criteria do we differentiate artists and art experts from other people?

Now we come back to the first question at the beginning: What is 'Art'? Taking a Wittgensteinian approach to the subject of aesthetics, such questions are rather trivial. Art is part of our forms of life, but it is not a single 'language game' as it were; it is a collection of practices connected by 'family resemblance' rather

than a shared essence. The collection of human activities that a language community agrees to call ‘art’ requires no sharp boundary dividing it from other life activities. “There are no conditions which are individually necessary and jointly sufficient for the application of these terms”—except, probably, necessary conditions that “art is a human activity and a work of art an artifact” (Glock, 33). Thus, the search for the essence of art should be abandoned. The issue with aesthetics is not that philosophers cannot give an adequate definition of aesthetic value but rather the feeling that such answers are needed.

Although a linguistic picture of aesthetics is reasonable on a descriptive level, I wish to further question whether our pragmatic use of the words ‘art’ and ‘beauty’ as umbrella terms is necessary or even desirable. My suggestion is rather negative. In our current age, artistic activities are still inseparable from wealth, power, and prestige. Claiming an activity as art, placing an object in an art gallery, or putting a piece up for auction all cast ‘spells’ to endow ‘art’ with real financial value. As I have previously argued, social, cultural, and political contexts play a large part in art appreciation. The way people view art has very acute realistic implications. When the word ‘art’ is already mystified to the degree that our community sees it under a sublime aura, we should be at least careful with our use of this word. Art historians’ decision to write about Duchamp’s *Fountain*, because it changed our ways of viewing art, is already a result of excluding some other activities as non-art or having less ‘aesthetic value.’ While art history does not have to claim it as art, the very act of discussing it makes it ‘worthy’ of attention. Labelling *Fountain* as ‘controversial’ does not change this dynamic—it still ensures that Duchamp’s work is studied, while other ‘non-art’ does not get the same treatment. This raises deeper questions about the politics of visibility in art. What determines whether something is ‘worth’ art-historical attention? Is it about innovation, institutional backing, media attention, or a combination of these?

This selective attention to ‘paradigm shifts’ in art history, such as Duchamp’s *Fountain*, does not necessarily depend on an artwork’s sophistication of artistic skills, the degree of pleasure it gives people, or its impact on people’s lives. Otherwise, Leni Riefenstahl’s *Triumph of the Will*, the infamous Nazi propaganda film, would be emphasized more in art history. Yes, film schools might study the techniques used in propaganda films, but compared to Duchamp’s *Fountain*, they are less known or emphasized in the public education of art. People rarely claim that propaganda films changed art history despite their far more

significant historical impact. The way art history is traditionally written is very different from regular history. The record of historical events is usually separated from their value judgments, but in art history, good works are remembered, and bad ones are ignored. Or it could be the other way around, the gesture of being included and emphasized in art history not only affirms but also endows a work with artistic value: if a piece of work is included in art history, then it must be 'good,' and if no one talks about it, it must be 'bad.' In that way, art history works as a self-reinforcing system: once an artwork is recognized as "historically important," its importance grows simply because it remains in discourse. Standard historiography, however, requires a more objective measure of events' impact on people's lives, whether good or bad, in deciding what to include. History will not write about what type of bread I ate this morning, even though the scene could be viewed as very aesthetic through a certain lens of art. Yet, history does not ignore World War II because of its horrifying massacres—on the contrary, it is part of the reasons it should be recorded and emphasized.

Maybe the question is: why do we only appreciate art as if we tell ourselves before stepping into a gallery, "Starting from this point on, I am looking at Art"? If ordinary human activities are often sanctified as art, what forbids us from adopting a similar attitude when viewing life itself? Why do we only appreciate John Cage's 4'33" when it is presented to us? Have we not experienced time or the absence of sound? Or has it just never occurred to us to pay attention to it outside the context of art appreciation? If life deserves to be treated seriously, then we should strive to live every part of it with care and attention. When we enter an environment that signals we are in the presence of 'art,' we are compelled to experience it with a certain reverence. However, when we dance for fun, we do not think, "This will be art"—even if it is full of expressiveness and rhythm—we dance because it is a part of living, something we do without needing to justify or elevate it. The reverence we reserve for art could just as well be extended to the so-called ordinary. But if we begin calling everything art and everyone an artist, the terms lose their meaning. Life has a variety and broadness above and beyond what we now count as 'aesthetic activity.' As Ernst Gombrich writes, "There really is no such thing as Art" (15); there only exists a variety of human activities in different times and places—people who paint, who sing, who pluck strings, people who build houses, who tell stories, who write poetry, and other countless events that enrich their own lives and the lives of others.

This is the irony of aesthetics: the less one cares about the supposed boundary between art and the ordinary, the more one perceives the richness of everyday life experience; the less one uses the word ‘beautiful,’ the more one can recognize the nuance of their feelings about something; the less we talk about aesthetics, the more we do, create, and entertain. As people document artists’ intentions, cultural contexts, and the impact of their works, the unique ‘aesthetic value’ loses its charm in history. The transient late-19th-century movement known as Aestheticism, centred on the doctrine that art exists for the sake of its formal beauty alone, soon turned into hollowness and kitsch. Aesthetic narratives not only disguise ideology and confuse us with its value judgments but also narrow diverse human experiences into a single dimension and trick us into thinking that life and the world are remote and incomprehensible.

So, what should we do? I cannot change our community’s use of aesthetic vocabulary or conception of art, as I myself use the word ‘beautiful’ as an interjection from time to time. But we can be aware of our use of aesthetic language, suspend our judgment, and be vigilant toward statements about what is art and what is not, what is the principle of beauty, and so on. And the rest is just to do—to listen, to read, to feel, to look, to create. As the artist Caryn Azoff says, “I asked myself, what was most important to me. It was first and foremost the act of doing” (Azoff).

Yet, as I decided to insert images next to my writing in this paper, I am inevitably adding meaning to these images, which perhaps does not exist in the first place. Readers should be aware of their potential bewitchment by my language, as there might be a charm in disenchantment, in deconstruction, in dissolving an issue instead of confronting it.

Works Cited

- Azoff, Caryn. www.carynazoff.com. Accessed 28 Mar. 2024.
- Berger, John. *Ways of Seeing*. The Schools, 1972.
- Gombrich, E. H. *The Story of Art*. Phaidon Press, 2016.
- Hans-Johann Glock. *A Wittgenstein Dictionary*. Oxford, Blackwell Publishing, 2007.
- The Museum of Modern Art. “Dada | MoMA.” The Museum of Modern Art, www.moma.org/collection/terms/dada/marcel-duchamp-and-the-ready-made.
- Wittgenstein, Ludwig. *Philosophical Investigation*. Macmillan, 1968.
- . *Tractatus Logico-Philosophicus*. 471st ed., Dover Publications, 1998.
- Wittgenstein, Ludwig, and Cyril Barrett. *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. University Of California Press, 2010.



Longtermism and Artificial Intelligence: An Analysis of Future Worlds

SABRINA SAVOIA

Developments in Artificial Intelligence (AI) present confusing moral implications for the advancement of humanity. This paper explores the risks to human happiness that AI can pose, detailing what we ought to do when technological advancements in AI hinge on the flourishing and well-being of future generations. I argue that we should not focus on maximizing the number of years future civilizations can have; rather, we should aim to provide the best quality of life for future people for as long as possible. Once the quantity of years to come outweighs the quality of years to come, we have failed morally, and we ought to allow human extinction.

Key words: longtermism, Artificial Intelligence (AI), ethics, population ethics, repugnant conclusion, philosophy of happiness, human flourishing, utility.

I. Introduction

Artificial Intelligence (AI) is evolving so rapidly that it is challenging to keep up with its ethical implications. Throughout this paper, I will propose three hypothetical worlds that may come into existence in light of the current trajectory of AI advancements. I argue that if we are burdened by the choice between a booming AI-dominant future with unhappy people and a future devoid of AI entirely with happy people, then we ought to choose the latter. I explore population ethics and human happiness, using AI as a possible antagonist to achievable human happiness. I argue that we ought to choose not to pursue AI-dominated futures if they compromise human happiness.

Imagine three potential futures: World A, World B, and World C. World A has a booming economy and a rapidly increasing population aided by AI but unhappy residents. World B is completely dominated by immortal AI that will effectively replace humans in a few centuries. World C is devoid entirely of AI, allowing human extinction once human beings are no longer able to live happy lives.

Though none of these futures are particularly appealing, at face value, World A seems like the best option. Yes, its residents are unhappy, but there are astronomically more people in World A than Worlds B or C, and most would argue that the quantity of years to come should surely compensate for lower quality of life when the numbers are so vast. The immortality of World B seems appealing as well, but it becomes ethically ambiguous when we consider the subservient role that humans will inevitably embody as a result. World C seems like a poor option. Why should we allow human extinction when there are alternative worlds that will encourage human flourishing for several more centuries?

In this paper, I will argue that among the three options, allowing World C to come about is the moral decision because World A and World B violate the principled moral obligations that we owe to future generations, as outlined by the philosophical theory of contractualism. I will provide a detailed overview of each world in Part II and offer my argument in favour of World C in Part III. In Part IV, I will consider whether I have given you a false choice, countering that my argument is flawed because it pessimistically assumes that the potential risk AI-dominated futures pose to human happiness will inevitably materialize. I will present William MacAskill's argument for optimism to defend this claim. I will then propose, and argue in favour of, World B', an immortal future that is dominated by

AI while still prioritizing human happiness, instead of World C. World B' is not meant to propose some fourth option. Instead, I am using World B' to demonstrate that the technological advancements in World B do not necessarily have to come at the cost of human happiness. Finally, I will argue in Part V that we should not be optimistic about the likelihood of World B', since we cannot be certain that it will come to fruition. I will conclude that allowing World C to transpire is still the moral option.

II. Overview

Consider World A, which presents a variation of Parfit's Repugnant Conclusion (Arrhenius). With the aid of AI technology, this society has soaring economic growth, with its population growing exponentially as the years progress. AI will help this society advance indefinitely, with its population grossly outnumbering that of our own by tens of billions. However, on a scale of well-being from 0 (the worst possible misery) to 10 (the most desirable quality of life), the aggregate well-being of this world's citizens stays at an unchanging level of 2. The inhabitants of this world will find themselves significantly less happy than they might have been had it not progressed so.

Next let us consider World B, which will outline a *slightly less* repugnant world than World A, encompassing a future of complete AI domination. Suppose in this world AI embodies human abilities and assumes full responsibility of human endeavours, replacing humans. The AI has reached sentience and is capable of a range of emotions that we would classify as human, possessing an aggregate well-being of 5. In this reality, however, humans still exist. The people in this world are oppressed by their more powerful AI counterparts, with their levels of aggregate well-being beginning at a level of 3 and decreasing steadily until it reaches a level of 0 and our species meets its end. This world will achieve full immortality through AI technological advancements, but this will happen at the expense of human happiness and well-being.

Finally, suppose World C comes about. This future is devoid of AI completely. In this world, humans advance at the rate which we are currently advancing, reaching an eventual population that has only grown by a dozen more billion. This world's population is considerably smaller than that of World A, and it lacks the sophisticated AI technology of World B, but its residents have an aggregate well-being of 10. Free from the advanced technology that would prolong human

existence, this world will only flourish as long as humans permit its evolution. It will meet its inevitable demise when humans can no longer advance in a manner that is conducive to their flourishing, eventually allowing human extinction to occur after a few more centuries.

III. Argument

Now that we understand what three potential futures can look like, we can assess the morality of each world. I argue that we should not solely focus on maximizing the number of years future civilizations can have; rather, we should aim to provide the best quality of life for future people for as long as possible. That is, we ought to value human happiness and well-being over every other factor. My argument goes as follows:

(P1) There are three potential future outcomes: World A, World B, and World C.

(P2) Permitting a potential future to advance wherein the quantity of years to come is prioritized over the quality of years to come is immoral.

(P3) World A and World B permit the advancement of futures with more years to come but lower levels of human happiness than World C.

(C) Allowing human extinction (World C) to transpire is more ethical than allowing an immortal future that is dominated by Artificial Intelligence (World A or World B) to flourish.

P1 is a morally neutral claim about potential futures. P2, however, holds a significant amount of weight in my argument, stating that prioritizing a future's quantity of life over its quality of life is immoral. P3 offers an analysis of the three worlds; World A and World B consist of larger populations that will advance for multiple more centuries, but their levels of aggregate well-being are significantly lower than they otherwise could be. Conversely, World C has flourishing levels of aggregate well-being but far fewer people than Worlds B and C. Since World C is the only world among the three that prioritizes quality of life over quantity of life, it is the most moral outcome according to P2, and we ought to allow it to come to fruition.

According to P2, we ought to prioritize the quality of years to come over the quantity of years to come. We are faced with two opposing views: utilitarianism and contractualism. Utilitarianism prioritizes actions that maximize overall happiness, or utility, for the greatest number of people, while contractualism focuses on moral principles that individuals would rationally agree upon in a hypothetical social contract.

A utilitarian objection to my argument could posit that the long-term benefits of vast population growth outweigh that of greater levels of aggregate well-being. They may take the stance that a more populated future will lead to immense technological progress and economic growth that could ultimately enhance overall utility for future generations. According to these principles, a utilitarian may defend the occurrence of Worlds A and B over World C because far more people would be brought into existence in the former worlds. I argue that the former utilitarian defense of the Repugnant Conclusion fails because it does not adequately respect the separateness of people.

Richard Yetter Chappell explores this stance, which rejects utilitarianism on the grounds that it treats people as mere receptacles of value. This anti-utilitarian stance argues that prioritizing maximal utility may promote an outcome's overall value, but it fails to consider the intrinsic value and agency of individuals (Chappell, 2). Utilitarianism poses the risk of treating people as 'value receptacles' when their value is solely measured according to their capacity to contribute to a desired outcome (Chappell, 2). When we value the future itself over the value of the people within the potential future, we are treating people as means to an end instead of treating them as ends in themselves.

Bernard Williams also touches upon P2, arguing that human prejudice often oversaturates our judgments about morality. Williams critiques the notion that philosophy should strive for objectivity and detachment from human concerns, asserting that this ideal of objectivity often leads philosophers to overlook the importance of human values (139). His argument can be applied to this 'dilemma of worlds' that we are facing, for we can use his stance that we ought to prioritize human values when we assess the morality of our three potential futures.

Choosing to bring about a future with a higher population over a future with a lower population seems better at face value. It seems that an increase in the number of people ought to result in greater positive outcomes. Though this

stance seems intuitively correct, I believe it is riddled with human prejudice. How can we justify choosing a world that allows human suffering over a world that promotes human flourishing, simply because the population size in the former exceeds that of the latter? When we approach this problem solely with the objectivity of numbers, we overlook the importance of human happiness. If we decide to allow World A or World B to advance over World C, subjective values such as happiness and positive well-being are outweighed by objective values such as population size. I argue that this is the problem with the utilitarian approach, and that we should look to contractualism instead.

Larry Temkin's *Aggregation within Lives* details a contractualist approach to the argument in P2. Temkin argues that traditional utilitarianism fails to adequately account for the complexity of moral decisions when considering the distribution of well-being within people's lives (3). Within utilitarianism's traditional understanding of aggregate well-being, distribution of well-being does not factor into its calculus. Temkin argues that the way in which the good is distributed must be considered when we assess the value of potential outcomes (10). Temkin posits that if we are burdened with choosing between one type of benefit that is of low quality and another type that is of high quality, a rational agent would posit that it is better for a smaller group to receive the higher quality benefit than for a larger number of people to receive the lower quality benefit (10).

Among our three options, utilitarian ethics favours World A or B: those worlds which promise to maximize longevity. However, they achieve this by maximizing their populations, rather than the level of well-being that each resident possesses. In contrast, World C favours the contractualist approach over the utilitarian approach, prioritizing the happiness of its citizens over the advancement of the world's population and society. According to Temkin's stance on trade-offs between quality and number, we ought to prioritize quality over quantity and allow World C to come about instead of World A or World B (14).

Temkin's Lollipops for Life Case further supports this stance. The case proposes two worlds: in one, countless people would each receive many licks of different lollipops over the course of their lives, but one innocent person would suffer unbearable agony for fifty years. In the second universe, each of the countless people would receive one less lick of a lollipop, but the innocent person would be spared agony and live a full, rich life (4-5). Much like the first variation of the Lollipops for Life Case, Worlds A and B propose futures that seem

desirable at face value but ultimately fail to respect the separateness of persons. According to Temkin, we ought to respect people and treat them as ends in themselves rather than mere receptacles of value. World C is the only option among the three that achieves this.

IV. Objection

The argument presented in Part III is flawed because it pessimistically operates under the assumption that the negative outcomes of World A or World B are inevitable. Why should we assume that population growth with the aid of AI technology will inevitably lead to a decline of human happiness and well-being? William MacAskill addresses this tendency to focus on the worst-case scenario of outcomes. MacAskill proposes an argument for optimism, which suggests that people frequently tend to produce good things for the sake of their being good and rarely tend to produce things for the sake of their being bad (218). It is pessimistic to assume that World A or World B are inevitable. We should be optimistic and assume that AI-dominant futures that do not compromise human well-being are possible.

Let us consider World B', another world that is dominated by AI, but which prioritizes the happiness of its human residents. It goes as follows: AI embodies human abilities and assumes full responsibility for human endeavours, effectively adopting a human experience alongside actual human beings. The AI has reached sentience and is capable of a range of emotions that we would classify as human, possessing an aggregate well-being of 7, opposing World B's aggregate well-being of 5. Humans in this world work alongside their AI counterparts, with their levels of aggregate well-being also staying at a level of 7, until humans face natural extinction. AI in this world will not face mortality after human extinction because AI can advance eternally. This world will achieve full immortality through technological advancements in AI, and the human beings within this world will possess a satisfactory level of aggregate well-being.

Like World B, World B' is dominated by AI; however, World B' differs because this domination does not transpire at the expense of human happiness and well-being. In World B', human levels of aggregate well-being are significantly higher than in World B, for the people stay at a level 7 instead of a rapidly decreasing level below 3. This modification to World B makes it less repugnant. Now we can ask: How does World B' compare to World C?

The primary argument in Part III focuses on the stance that Worlds A and B violate the moral principles that we owe to future generations as outlined by contractualism, but this argument crumbles when we consider World B' because humans are no longer at such a low level of aggregate well-being that their lives are not worth living. P2 states that permitting a potential future to advance wherein the quantity of years to come is held to a higher standard than the quality of years to come is immoral. World B' cannot be classified as a future that prioritizes quantity over quality in principle, since its sentient beings possess a decent level of aggregate well-being.

Allowing World B' to advance seems to have similar moral implications to the morality of allowing World C to advance. World B' will advance indefinitely, and its human residents' happiness levels will stay acceptable. Alternatively, World C will advance for a fraction of the time that World B' will, and have far fewer sentient beings, but its residents will possess an exceptional level of happiness at 10. It seems as though the amount of human suffering in World B' is not sufficient to justify human extinction (World C) coming about in its place. We can and *should* assume that World A and World B will not reach the worst of their negative outcomes because of MacAskill's argument for optimism, hoping that World B' will come about instead of World B.

V. Response

Although Part IV correctly asserts that World B' is superior to both World A and World B, the claim that World B' and World C share identical moral implications is incorrect. This is because Part IV's argument for optimism overlooks the crucial role that probability discounting plays when evaluating future outcomes.

Probability discounting refers to the tendency to devalue rewards or outcomes as their probability of occurrence decreases. Bradley Monton argues against the practice of probability discounting in orthodox decision theory. He uses the thought experiment of Pascal's Mugging (Bostrom) to depict this kind of irrational decision-making. This thought experiment, proposed by philosopher Nick Bostrom, presents a scenario wherein a stranger promises to give you a billion dollars if you give them the small amount of money in your wallet. This case highlights how unintuitive it seems to risk choosing an outcome with a negligible probability of occurrence even though it has an astronomically large reward.

Monton argues that traditional principles of discounting and maximizing expected utility may lead to counterintuitive results in cases like Pascal's Mugging, where the probability of an event is extremely low, but its potential payoffs are extremely high (12-13).

I agree with Monton's stance that risking negative outcomes in cases where the consequences are severe but highly improbable is counterintuitive, and I will apply it to the dilemma of worlds that we faced prior with a case called *The Pill*.

The Pill: You have a moderately happy life, with your well-being sitting at a comfortable place of 6 on a scale from 0 to 10. You are approached by a man on the street one day who offers you two identical pills. The man says that one pill will transport you to an utter utopia, while the other will transport you to a world that is riddled with pain and suffering. Once you decide to take a pill, your decision cannot be undone¹. The pills look identical, so you have an equal chance of choosing either eternal agony or eternal ecstasy. You have two options: take a pill and risk being transported to a world full of suffering or politely decline the man's offer and continue with your day.

In the case of *The Pill*, going about one's day is the same as allowing World C (human extinction) to occur, while deciding to take a pill is synonymous with opting to risk the chance of the other AI-dominated worlds to transpire. It seems *prima facie* wrong to decide to take the pill. Your levels of aggregate well-being are at a satisfactory level of 6—why risk a 50 percent chance of living in eternal agony?

For the sake of MacAskill's argument, let us assume that the 50/50 probability of choosing between agony and ecstasy is pessimistic, and that the probability is closer to 99/1. Instead of proposing two pills that have equal probabilities of transporting you to these different worlds, the man presents you with 100 pills. He tells you that 99 of the pills will transport you to a utopia, and 1 pill will transport you to a dystopia. It seems that even risking a negative outcome with a smaller probability of happening is counterintuitive. Why would we risk any chance of

¹ This aspect of *the Pill* case is meant to mirror MacAskill's theory of value lock-in, which refers to the idea that once society decides upon certain values or norms, it becomes increasingly difficult to change or deviate from them, even if they are no longer desirable.

ending up in a world of endless suffering?

Finally, let us examine a scenario that entails a negative outcome with a miniscule probability of occurrence. Allow me to propose one final case of *The Pill*, but instead of the probability of utopia vs. anti-utopia being 50/50 or 99/1, the man assures you that the probability of choosing the pill that transports you to utopia is 99.999/0.001. The probability of choosing the ecstasy pill now grossly outweighs the probability of taking the pill with dire consequences. Does this change how we should act if we are faced with the dilemma of *The Pill*?

Monton argues that we ought to avoid maximizing expected utility and refrain from discounting low probabilities to zero—a stance that I wholeheartedly agree with. Just because it is extremely unlikely that the pill will transport you to the dystopia does not mean that it is *impossible*. There is still an immense risk involved when we opt to take the pill, even if the probability of such risk occurring is less than 1 percent. The mere potential for the negative outcome to happen outweighs the reality of its extraordinarily low probability.

There is always a risk involved when considering AI technology aiding the continuation of civilization. World A and World B pose the highest risk to human happiness, for they entail potential futures wherein aggregate levels of well-being are dangerously low. World B' poses a significantly lower risk, only exposing human beings to the potential of human happiness at level 7. There is only one world, however, that poses no risk to human happiness: World C. Given that World C is the world that preserves the highest quality of human happiness, it is the world that we ought to choose.

VI. Conclusion

Advancements in AI come with grave uncertainty. An exploration of three potential futures—World A, World B, World C—has argued that opting for World C is the most moral decision among the three options based on the violation of moral obligations to future generations found in Worlds A and B, as outlined by contractualism. Despite considering alternative scenarios, such as World B', the fact that we cannot be certain whether it will ever occur undermines the optimism associated with it. The case of *The Pill* further argues that risking a negative outcome with a miniscule probability of occurring is counterintuitive, concluding that we ought to allow World C—or human extinction—to transpire.

Works Cited

- Arrhenius, Gustaf. "Gustaf Arrhenius, the Very Repugnant Conclusion." *PhilPapers*, 1 Jan. 1970, philpapers.org/rec/ARRTVR.
- Bostrom, N. "Pascal's mugging." *Analysis*, vol. 69, no. 3, 14 June 2009, pp. 443–445, <https://doi.org/10.1093/analys/anp062>.
- Chappell, Richard Yetter. "Value receptacles." *Nous*, vol. 49, no. 2, 15 Apr. 2013, pp. 322–332, <https://doi.org/10.1111/nous.12023>.
- MacAskill, William. *What We Owe the Future*. Basic Books, 2023.
- Monton, Bradley. "Bradley Monton, How to Avoid Maximizing Expected Utility." *PhilPapers*, 1 Jan. 1970, philpapers.org/rec/MONHTA-2.
- Temkin, Larry S. "Aggregation within lives." *Social Philosophy and Policy*, vol. 26, no. 1, 24 Nov. 2008, pp. 1–29, <https://doi.org/10.1017/s0265052509090013>.
- Williams, Bernard. "Bernard Williams: Complete philosophical publications." *Philosophy as Humanistic Discipline*, 31 Dec. 2009, pp. 215–227, <https://doi.org/10.1515/9781400827091.215>



The Inherent Strangeness of Ideal Coherence

LINDSAY COBB

Ideally Coherent Eccentrics (ICEs)—people who have perfectly coherent worldviews but unheard-of normative values—are generally taken to provide strong arguments in favour of attitude-independent theories of normativity. In her 2009 paper *In Defense of Future Tuesday Indifference* philosopher Sharon Street argues that several common examples of ICEs are not irrational in the way they initially seem to be. This paper seeks to expand and generalize her argument, showing that it is the common feature between all ICEs—ideal coherence itself—which makes them fail to provide strong arguments in favour of attitude-independent theories of normativity. This result has important ramifications for metaethical debate and strengthens the case in favour of attitude-dependent theories of normativity.

Keywords: Ideally Coherent Eccentrics, rationality, metaethics, normativity, Alan Gibbard, Sharon Street

In her 2009 paper *In Defense of Future Tuesday Indifference*, the philosopher Sharon Street argues that “Ideally Coherent Eccentrics” (ICEs)—people who have perfectly coherent worldviews but unheard-of normative values—aren’t irrational for valuing the strange things they value. Street argues this by going through three common examples of ICEs and explaining how the powerful intuitions that their values are intrinsically irrational, which these examples are meant to generate, don’t hold up under a more careful description of each of the cases. I will try to generalize Street’s point by arguing that these cases fail to generate their target intuitions for a common reason—the inherent strangeness of ideal coherence itself—and, therefore, that it is not ICEs’ peculiar views that make them eccentric, but rather their very coherence. I will then suggest as a conclusion that ICEs—generally conceived—cease to be a challenge to attitude-dependent normative theories.

Conceptual Background

In order to understand what follows, a few definitions can be illuminating:

- **Ideal Coherence:** To be ideally coherent, an agent must meet two criteria. First, her values must be completely consistent, never contradicting themselves. Second, she must make no mistakes of practical reason or factual mistakes. The agent thus has perfect knowledge of the natural facts and is always completely rational, believing and doing everything entailed by her values.
- **Evaluative Attitude:** An evaluative attitude is a value-judgement held to be true by an agent. When an agent says that some thing ‘x’ is ‘good,’ or ‘bad,’ ‘beautiful’ or ‘ugly,’ ‘better’ or ‘worse,’ and so on, she has an evaluative attitude regarding x.
- **Prudential Reason:** Prudential reason refers to any “self-regarding” reason. This includes matters of practical rationality and matters of fact (insofar as agents take themselves to have reasons to believe the truth). It is a mistake of prudential reason to be wrong by your own account, contradict your own values and desires, or to make factual mistakes if you take the facts to matter. This also applies to related phrases like ‘prudential mistake.’

Why ICEs Matter

The reason Street and I are concerned with ICEs is that they are important in discerning between attitude-dependent and attitude-independent theories of normativity. Attitude-dependent theories of normativity hold that there are no normative facts which hold independently of one's evaluative attitudes that provide an agent with reasons about how they must live. In contrast, attitude-independent theories of normativity hold that there are such normative facts. In other words, attitude-independent theories hold that "some values and desires are intrinsically irrational" (Street, 274) while attitude-dependent theories deny this.

In the case of actual human beings, these theories tend toward similar conclusions because most humans tend to hold somewhat similar evaluative attitudes regarding basic issues of morality. For instance, almost all real people would agree that pain is bad, or that killing people for no reason is wrong. ICEs allow us to isolate cases where these theories of normativity differ; in other words, cases where an agent is making no prudential mistakes but has normative views the attitude-independent theorist would describe as 'intrinsically irrational'. In such cases, the attitude-dependent theorist is forced to say that such agents are rational, while the attitude-independent theorist must say the opposite.

Prima facie, ICEs seem to support attitude-independent normative theories, as some of them would force the attitude-dependent normative theorist to adopt seemingly strange and intuitively wrong positions regarding the validity of certain moral claims. Three famous examples are relevant here. First, imagine an ideally coherent anorexic who values nothing else other than thinness, even if it means starving herself to death. Second, imagine an ideally coherent woman with *Future Tuesday Indifference* (FTI), who simply does not care what happens to her on future Tuesdays and will prefer agony on any Tuesday to a small pain on any other day. Third, imagine an ideally coherent Caligula who loves to cause suffering and thus believes he must always cause the greatest possible suffering to others.

In each instance the attitude-dependent normative theorist is forced to commit herself to an unintuitive result: the anorexic really does have most normative reason to starve herself, the woman with FTI really does have most normative reason to prefer agony on Tuesday to slight discomfort on Wednesday, and Caligula really does have most normative reason to hurt people.

However, once these examples are clarified, the position of the atti-

tude-dependent theorist is improved dramatically. A proper understanding of each of these examples will reveal that ICEs fail to generate the intuition that their values are intrinsically irrational. I will demonstrate this by going through the first example—the anorexic—in detail, before briefly explaining how Street applies similar reasoning to the other two examples of ICEs.

Clarifying the Examples

The first ICE example Street deals with is the ideally coherent anorexic, an individual who wants to starve herself so as to be as skinny as possible, and takes this as her highest-order value even if it entails her dying of starvation. Street rightly claims that the powerful intuitions generated by this example are “smuggled in” from our normal reactions to the case of real-life anorexics (277). In the case of real-life anorexia, most of us would say to anyone suffering from the condition that they were wrong in both their actions and beliefs, that they are being irrational, that they should eat in the interest of their lives, other values¹, and so on. We would be right to say this about real-world anorexics because they *are* irrational. Real-world anorexics are deluded, as they believe false or irrational propositions, such as “starving myself is the only way I can be beautiful,” “I’m fat” (despite actually being dangerously thin), “starving myself is how I may have control over my life,” and so on (Dalhoff et al.). Indeed, not only do actual anorexics make prudential mistakes, they also suffer physically, psychologically and socially greatly as a result; (Waldman et al.)². A majority of them become extremely unhappy while their other goals and values—their hopes for their life, their care for their family or friends, and so on—are harmed in the process of starving themselves.

All of this is utterly alien to the ideally coherent anorexic, who is—by stipulation—not deluded in any way. The ideally coherent anorexic does not hold the false belief that she is fat, nor the false belief that thinness is necessary for beauty, nor is she motivated by an irrational need for control over her weight. Instead, the ideally coherent anorexic must be conceptualized as taking thinness itself as an *inherently valuable good*, so valuable that it eclipses all other values and is worth dying for. Unlike real women, the ideally coherent anorexic doesn’t care about anything else that contradicts the value of thinness and can only be

¹ In the case of real-world anorexics, they hold values which are not all related to their being thin.

² Where anorexia is *defined* by cognitive distortions and mistakes of prudential reason.

made happy by dying in the pursuit of being thin. Thinness here does not serve any further purpose than simply being thin, understood as being the supreme good, all other potential goods becoming insignificant compared to it.

As should be obvious, the ideally coherent anorexic displays significant differences with real-life anorexics. Real-life anorexics—like all other actual humans—are not ideally coherent and make mistakes of reason as a result of their anorexia. Indeed, what we usually mean by the word ‘anorexia’ is a specific delusion or rational mistake (Street 277; Waldman et al.)! Instead, she is best imagined as a kind of strange “alien visitor” (Street, 281) about whom our intuitions are *much* less powerful. Imagine a scenario where alien creatures came to Earth and told us that starving themselves to death was the only thing they cared about and the only way for them to be happy—whilst absolutely convincing us that they made no rational mistakes nor factual errors in arriving at this conclusion. Would we be able to have a clear and powerful intuitive response that they were wrong? Or, at least, would we be able to have such a response as strongly as we would in the case of a human suffering from actual anorexia? Certainly not, as their displaying fundamental differences from human beings (biologically, culturally, contextually, etc.) would likely lead us to suspend the beliefs and assumptions and beliefs which underlie and provide the foundation of our moral deliberations regarding human beings and other earthly creatures.

Similarly, the ideally coherent woman with FTI who would prefer agony on Tuesday to slight discomfort on Wednesday would need to have a phenomenology³ so alien to our usual experience that it is best thought of as *literally* alien. Street explains that—like the anorexic—the woman with FTI is completely dissimilar to real human beings, and it is our instincts about real people that generate our intuitions according to which the woman with FTI is wrong and irrational to prefer full agony on a Tuesday to a papercut on a Wednesday.

If your friend said they didn’t care about what happened to them on Tuesday, you would likely respond that they were being irrational, but this is because your friend could only come to believe this by making a prudential mistake, such as the belief that they were someone else on Tuesday. The ideally coherent woman with FTI is—by stipulation—making no such mistake; she rather seems

³The notion of phenomenology here should not be taken as a reference to any particular theorist—Husserl, Heidegger, Merleau-Ponty, and so on—but instead as a word for the full content of our experience, including its pre-reflective mental content not usually captured by the word ‘experience.’ See: Charles Taylor’s use of the term for similar uses.

intrinsically programmed not to care. Street asks us to imagine the same thought experiment as before: to think of this woman as a member of an alien species who—for some evolutionary reason—have evolved to think this way. Here too, our strong initial intuition about the wrongness of the agent's beliefs now seems unjustified, as it would seem incoherent to suggest that an alien creature is simply wrong or irrational to not care about Tuesdays due to their evolved programming, but that we are right in following our evolved emotional programming (such as to care about our children, or to want to be admired by others) (Street 287).

The next case is the case of someone with a dual condition: they have FTI but also *Present Tuesday Horror*. This is someone who doesn't care about future Tuesdays but cares about what happens to them on a Tuesday from the moment Tuesday starts. It may not be obvious at first glance, but this person is the same as the previously mentioned Caligula. This is because her error—if she is making one—is a *moral* error rather than a *prudential* error. She has an appropriate understanding of the facts: she completely understands that she is causing herself to suffer horribly on these future Tuesdays and that she is the same person on Tuesday as any other day. Her only 'error' is to place an insufficient or incorrect amount of moral value on suffering. This is like Caligula—who perfectly understands the pain he causes—and remains unmoved by his victims' pleas for help.

Let us finally turn to the case of Caligula. In the case of Caligula, there is an important distinction between something being morally required and something giving an agent normative reason to act. It is fully coherent for the attitude-dependent theorist to claim that Caligula is not irrational but that she disapproves of his actions anyway, that he is immoral, and that she has moral reasons to try and stop him (Street, 292). Importantly, it is not obvious that something being a requirement of morality necessarily gives any agent reason to do it, especially if she has strong reasons to do the opposite. When pressed, Caligula would argue that he does indeed have strong reasons to do the opposite of what is required by morality, since causing harm is the only thing he cares for and about. The question, therefore, is not whether torturing people is immoral, but rather whether Caligula has a normative reason to be moral at all (292). In other words, our question is whether or not Caligula is irrational for not thinking he has a valid normative reason to be moral or act in a moral way.

Here too, it seems best not to think of Caligula as human, as doing so

‘smuggles in’ our intuitions about evil humans, who are almost always making a prudential mistake, such as falsely believing their victims aren’t fully human⁴ or contradicting by their action a value they claim to hold⁵. If we conceive of Caligula as a hyper-advanced AI system with a value function that takes suffering to be the only good, the example becomes clearer. In this case, it’s clear that this AI system has a value function we think is morally wrong, but it’s less clear that such an AI system is, therefore, irrational to act on this value function. I think the intuition here is certainly less clear than our initial reaction to the idea of a human torturer, whom we intuitively regard as irrational.

One ICE to Rule Them All

In each of the previous cases, once the full strangeness of the ICE is grasped, the initial intuitions we might have about them seem to disappear, weaken, shift, or become confused. Simply diffusing these three examples doesn’t bolster the attitude-dependent theorists’ case very much; she needs to show that *all* ICEs will fail to show that attitude-dependent views seem implausible. I think it possible for the attitude-dependent theorist to produce such an argument by showing that it is the common factor between *all* ICEs—ideal coherence—that causes the loss of intuition-generating potential.

To illustrate this point, imagine one final ICE: the case of Moral Mildred. Moral Mildred is ideally coherent, and holds ethical normative views that align with whatever the attitude-independent theorist would endorse as correct. Since Moral Mildred is completely coherent, she has no other normative views that conflict with these in any way whatsoever, and all of her higher order values tell her to accept the moral norms derived from them as her lower order norms. Effectively, this means Moral Mildred has no other values beyond her moral values, as any other value (for instance, rooting for a particular sports team, loving a particular person, and so on) could potentially conflict with her universally applicable moral values.

There are three important things to realize about Moral Mildred:

1. Since her moral views are fairly standard, the only *truly* strange thing about her

⁴Think about Himmler and the Final Solution as an example.

⁵Think about cigarettes as an example. I continue to smoke even though I say I value not getting lung cancer and that I know that cigarettes cause lung cancer. I am practically irrational because my actions contradict my own values.

is that she is ideally coherent.

Under the attitude-independent theory of normativity, Moral Mildred should motivate within the observer the strong intuition that she is perfectly reasonable, and much more rational than Caligula, the woman with FTI, and the anorexic. Indeed, an attitude-independent theory of normativity entails that she is even more rational than any of us. The only sense in which Moral Mildred would seem extreme for the attitude-independent theorist is that she would seem *extremely correct*. However, an attitude-dependent theory of normativity would predict that Moral Mildred will seem intuitively similar to the other ICEs; that she would seem strange and alien in a way that confuses our intuitions and seems baffling to normal people. In this way, the case of Moral Mildred allows us to isolate the source of our intuitions that ICEs are irrational. Indeed, if these intuitions came from a recognition that the odd normative views of ICEs are normatively and attitude-independently wrong, then Moral Mildred should seem extremely dissimilar from them—as her views are accepted as right. On the other hand, if the intuitions only come from “smuggling in” our intuitions about real people, Mildred should seem similar to other ICEs.

2. Moral Mildred is impossible and very dissimilar to actual examples of every moral people.

Although some actual people strongly and strictly adhere to their moral standards—for example, ascetic monks or utilitarians who give large amounts to charity and only keep what is strictly necessary to their livelihood—they also have other normative views, which sometimes conflict with their morally normative views. These people make tremendous sacrifices for their moral views *precisely because they care about other things and choose to prioritize morality over these other things*. Furthermore, they experience uncertainty and conflict, as they wonder if they are right about the non-normative facts, if one normative end is more important than another, and so on. In contrast, Moral Mildred would experience none of these emotions, uncertainties, or conflicts.

3. Moral Mildred experiences a very particular phenomenology.

- Firstly, she experiences her moral views—no matter what they are, Kantian, Utilitarian, and so on—as the only things of normative value, nothing else being normatively relevant to her.
- Secondly, she would be infinitely selfless, going to great lengths to do what she thought was morally right with no regard for any other concerns—including her own suffering. She would only care about preserving her life insofar as it was instrumentally useful to do what was right or was itself of moral concern⁶.
- Thirdly, she would not experience this selflessness or its consequences as upsetting, as she only cares about doing what's right. Given the choice between not being able to help anybody whilst avoiding suffering and being able to help many people but suffering tremendous torture, Moral Mildred would be flooded with relief upon hearing the second option, choosing it in a heartbeat. Moral Mildred would readily and *happily* agree to be tortured forever if it meant saving others from the same fate, experiencing this agreement as being a purely good thing since it furthers the only normative end she cares about: doing what is right.

When described this way, does Moral Mildred seem any more rational than the anorexic? Or the woman with FTI? Is the only thing that is extreme about her how rational and correct she is? I do not think so, and this is already a poor result for the attitude-independent theorist hoping to use ICEs as a knock-down argument against her attitude-dependent opponents. Contrary to what attitude-independent theories would predict, Moral Mildred does not seem perfectly rational but rather strange and unintuitive, similarly to other ICEs. However, it gets worse for the attitude-independent theorist, as Moral Mildred is not merely strange in a way that is similar to the other ICEs but in a way which is *very* similar to the woman with FTI. This gives us good reason to prefer the attitude-dependent account, which predicts exactly this sort of similarity.

Recall that the strangest part of the woman with FTI (and not *Present Tuesday Horror*) is that she experiences pain—in the fullest sense of the word, not merely the sensation but also a “hedonic disliking” of the sensation (Street,

6 If we imagine Moral Mildred as utilitarian for instance, her own suffering matters to her only in the exact same sense as someone else's suffering and she would endure any amount of suffering if it meant preventing more suffering for others. If we imagine her as Kantian then she wouldn't kill herself—as Korsgaard rightly explains in *The Sources of Normativity* that suicide is wrong in a Kantian framework—but in taking others as her only ends she would be willing to endure extreme things to help them.

290)—on Tuesdays, but simply isn’t motivated to care. The woman with FTI experiences agony on a Tuesday as intensely as a normal person and her body is in “all out revolt” (290) against that agony, but at her core emotional being, she experiences complete calm. Moral Mildred would be in exactly the same situation should she be forced to go through tremendous torture in a case where this would be the best moral option. She would experience intensely painful sensations, and equally intense hedonic dislikings of those sensations as the woman with FTI on Tuesdays, but her core emotional state would remain unbothered due to her accomplishing the only thing she cares about—doing what is right.

For the attitude-independent theorist to use ICEs as an argument against attitude-dependent theorists, Moral Mildred would have to seem much more rational than the woman with FTI. Given this comparison, it becomes clear that Moral Mildred seems just as irrational as the woman with FTI. Indeed, the woman with FTI *may* be more relatable than Moral Mildred, as Moral Mildred is not merely unbothered by this experience of extreme pain but actually sees it as *good and desirable* because she is actively achieving her normative ends.

Conclusion

Sharon Street is right to say that common examples of ICEs used to argue against attitude-dependent theories fail to generate the powerful intuitions they are meant to generate against such theories. This is because ideal coherence itself produces cases so strange that they are inherently unintuitive, even when the normative views that are stipulated to ideally cohere are utterly ordinary. Cases like Moral Mildred demonstrate that the unintuitiveness of ICEs comes not from their strange values—as the attitude-independent theories would predict—but instead from being ideally coherent. This makes ICEs fail in general as knock-down arguments against attitude-dependent theories as, once they are properly understood, they always fail to produce strong, clear intuitions regarding actual humans due to their ideal coherence.

Works Cited

- Street, Sharon. "IN DEFENSE of FUTURE TUESDAY INDIFFERENCE: IDEALLY COHERENT ECCENTRICS and the CONTINGENCY of WHAT MATTERS," *Philosophical Issues* 19, no. 1 October 2009. 273–98, <https://doi.org/10.1111/j.1533-6077.2009.00170.x>
- Dalhoff, Anke W.; Romero Frausto H.; Romer G.; Wessing I., "Perceptive Body Image Distortion in Adolescent Anorexia Nervosa: Changes after Treatment," *Frontiers in Psychiatry* 10 (October 15, 2019), <https://doi.org/10.3389/fpsyt.2019.00748>
- Korsgaard, Christine M. *The Sources of Normativity*. Cambridge Univ. Press, 2014.
- Waldman, Amanda; Loomes, Rachel; Mountford, Victoria A; Tchanturia, Kate, "Attitudinal and Perceptual Factors in Body Image Distortion: An Exploratory Study in Patients with Anorexia Nervosa," *Journal of Eating Disorders* 1, no. 1 (2013): 17, <https://doi.org/10.1186/2050-2974-1-17>



Veiling Illusions: Language and Reason as Veils Concealing the True Essence of the World

ANNA BISTOUR

This essay examines Friedrich Nietzsche's critique of language and reason, arguing that they act as veils concealing the chaotic essence of the world. Challenging German Idealism, particularly the philosophies of Immanuel Kant and Johann Gottlieb Fichte, Nietzsche dismantles the notion of reason as humanity's highest faculty. Indeed, German Idealists, including Fichte, viewed reason as the foundation for objective truths and the construction of the autonomous 'I.' In *Wissenschaftslehre*, Fichte positions the ego as the central source of certainty, yet Nietzsche critiques this framework, highlighting its paradox: the ego, shaped by reason, limits human freedom rather than enabling it. Language, deeply entwined with reason, is similarly critiqued as an arbitrary and reductive system that distorts reality by imposing structured classifications. Nietzsche contends that reason and language fail to capture the deeper truths of existence, creating illusions that confine human agency and obscure instinctual drives. This essay further argues that the overemphasis on reason in philosophical systems suppresses the natural instincts and desires necessary for human flourishing. Nietzsche contrasts the mechanistic worldview of German Idealism, including Fichte's egocentric system, with a call to embrace creativity and instinct as the true foundations of human nature. Ultimately, this essay advocates for rejecting the dominance of reason and language to foster a more authentic understanding of the self and the world.

Key words: Nietzsche, language, reason, German idealism, enlightenment, illusions, decadence, autonomy, chaos, instincts, rationality, self-determination, will, ego.

Introduction

In the wake of the European Enlightenment, which characterized the 17th and 18th centuries, the consensus was to place reason as the defining feature of human nature. Regarded as the highest human faculty, reason was deemed capable of organizing the disorderly world. Yet, Friedrich Nietzsche (1844-1900) devoted a significant part of his work to attacking and refuting the common belief concerning the preeminence of reason. In his work *Twilight of the Idols*, he underscores in the section “Reason in Philosophy” the connection between language and reason (Nietzsche, 21):

“Language belongs in its origin to the age of the most rudimentary form of psychology: we find ourselves in the midst of a rude fetishism when we call to mind the basic presuppositions of the metaphysics of language—which is to say, of reason. It is this which sees everywhere deed and doer; this which believes in will as cause in general; this which believes in the ‘ego’, in the ego as being, in the ego as substance, and which projects its belief in the ego-substance on to all things—only thus does it create the concept ‘thing’ ... Being is everywhere thought in, foisted on, as cause; it is only from the conception ‘ego’ that there follows, derivatively, the concept ‘being’ ... At the beginning stands the great fateful error that the will is something that produces an effect—that the will is a faculty...”

This passage sheds light on language’s dual function as both a source of illusion and as strengthening the misleading power of reason. Reason apprehends the world through classifications. Language, as a human creation entwined with reason, serves as a communicative tool to understand others, share ideas, and gather individuals, through the use of (artificial) classifications. In this paper, I argue that language and reason are not perfect tools to express the world and that, as fixed meaning-bearing instruments, they limit human self-determination and creative minds, for they fail to capture the deeper meaning of human perceptions. Challenging German Idealism, this paper delves into the paradoxical relationship between language and reason, demonstrating that an overemphasis on both these concepts can actually diminish human autonomy rather than enhance it.

Opposing Immanuel Kant's Prevailing Doxa

Nietzsche fundamentally rejects the idea of a single, ultimate truth constructed and attained by reason, in opposition to one of the most significant figures of the Enlightenment, Immanuel Kant. Writing in the midst of the Enlightenment, Kant aimed at unchaining humans from ignorance and error through advancements of knowledge, believing objective truths were the means of emancipation. In coherence with other German idealists, he suggested life must be understood through the prisms of reason as logic and critical thinking. Kant sought objective truths via a rigorous and transcendental system, based on unchanged and stable categories, aiming to allow the thinker to apprehend and organize the world methodologically. His system posited that certain facts belonged to the category of *a priori* knowledge—knowledge independent of experience—arising from the mind's innate structures, which organize and condition human experience (Kant, 24). Resting upon the use of categories, Kant's method encapsulates the capacities of the rational mind, including sensibility, understanding, and reasoning; the latter being the highest of them all as it is solely grounded in reason. In the previously cited excerpt from *The Twilight of the Idols*, Nietzsche attacks this belief. He rejects the idea of a transcendental rationality, arguing that when reason permeates all forms of thought or expression, it creates appearances and concepts empty of meaning. Prioritizing reason above all else and rationalizing all experiences and instincts seems rather unnatural in this way—a sign of *décadence* (Nietzsche, 6).

Reason and Language as Creating an Untruthful Representation of the World

Analyzing the origin of language, Nietzsche challenges the conception of language as a demonstration of reason, suggesting that it may be misleading, going so far as arguing that language may be involved in constructing a false reality. Nietzsche refers to language as “the most rudimentary form of psychology,” the simplest form of logic, heavily subject to improvement (Nietzsche, 21). Derived from the Greek *ψυχή*—*psychē*, meaning soul or the spirit—and *λόγος*—*lōgos*, logic or reason, the term ‘psychology’ refers to the scientific account of the mind. In introducing language as a ‘psychology,’ Nietzsche suggests that through discourse and speech, humans explain their psychological patterns, habits, and thoughts. As a means to elevate humans by providing them with a better understanding of themselves, each other, and the world, language mediates and conceals discrepancies between the arranged reason and the chaotic outside world.

Extending beyond words, language manifests itself, *inter alia*, through symbols, sign language, customs of gift-giving, religious rituals, art, music. Although language represents an optimal tool to understand the world, the rules of grammar erroneously subordinate it to the almighty ego.

Witnessing humans' attitude toward language, Nietzsche criticizes them, pointing out that humans partake in fetishistic activities based on the assumption that language possesses some kind of magical power or agency.¹ Nietzsche pushes us to discover ourselves in the middle of "a rude fetishism," (*Fetischwesen*) where we realize the influence of language upon us. However, this realization only occurs later, after we have come to understand the extent to which we are subjugated by the control of language (Nietzsche, 21). Nietzsche suggests that language embodies a kind of powerful spirit, an illusory authority which creates within individuals the sentiment that reason, as the ever-present precursor of language, allows us to conceive the world. Drawn from preconceived representations, words and concepts stabilize our world interpretations, where the reasonable subject—the doer, or 'I,' is differentiated from the distant and alien object—the deed. Rooted in reason, language serves to build a meaningful and stable world, disguising the chaos behind organized structures.

However, Nietzsche argues that this 'fetishism' and mode of thinking do not exemplify the world as it truly is: language too easily becomes an illusion, a flawed tool controlling humans' perceptions through predefined images. The misconception of language as a sense-maker of chaos perhaps arises from humans' high appraisal of language, probably stemming from the long-lasting belief that reason, humans' greatest faculty, supports language. Reason being as Nietzsche puts it "the metaphysics of language," plays an active role in defining the structure conditioning our perceptions of the world (21). In calling descriptive language a fallacious instrument to comprehend the world, Nietzsche contests humans' over-reliance on language, suggesting that reason and, by derivation, language in its most traditional view are flawed tools, and are thus unsuccessful in expressing the human experience.

The Limitations of Reason in Organizing Perception

The act of reasoning seems fundamentally grounded in language: rea-

¹ Nietzsche criticizes social expectations and consequently humans' behaviours, in their taking part as a group in fetishistic activities. He viewed the worshipping of language and reason as a manifestation of blind faith.

son apprehends reality through rigid categories shaped by words—manufactured linguistic frameworks. Therefore, if vocabulary is limited, humans’ apprehension of the world will be correspondingly bracketed. One can only think, express ideas, and apprehend the world with words and categories that one possesses—language thus restricts perspectives. Consequently, humans become subject to language, in that they must submit to a preconceived reality, which language has already created.

For instance, describing what he understands to be the behaviour of language in comparing meaning to the Sun and language to the planets orbiting it, Nietzsche claims that language turns itself around to go back to its origin, moving around the subject, but never fully reaching the expected aim (Nietzsche, 20). In other words, language fails to grasp the essence of a subject, as it never exactly reaches a perfect account of the object of discourse. Misguided by reason classifying the world under factual representations, language overlooks the true nature of a subject, categorizing it without truly paying attention to nuances within. Language then becomes completely arbitrary, simply arranging the world in predetermined categories that reason alone produced. It remains incapable of helping humans accurately interpret reality, only apprehending its surface, failing to grasp truthfully the essence of the world. Nietzsche uses this metaphor to criticize one of the most influential German Idealists, Georg Wilhelm Friedrich Hegel, who developed a philosophical system he referred to as the *Dialectic*. Rooted in logic, this system proffered to resolve contradictions through a circular process involving thesis, antithesis, and synthesis (Hegel). Hegel’s emphasis on the use of reason through language, reflected in the choice of the term ‘Dialectic’ — literally ‘the art of debate’ — underscores the centrality of language in German idealist philosophy, rejecting empiricism and favouring rationality in explaining experiences.

Whilst German Idealists maintain that reason is the absolute ground of objective truths, Nietzsche critiques the notion that reason reigns supreme among our faculties. German Idealists argue that human nature is distinguished from animals by their capacity for reasoning, which allows for analysis, planning, and anticipation. However, Nietzsche’s critique highlights the dangers of excessive reliance on reason, suggesting that it can lead to false assumptions. In categorizing the world, reason adheres to a pattern, constructing reality in a mechanistic fashion, but failing to capture the entire meaning of our existence. Nietzsche

goes so far as to personify reason, claiming that it “sees,” “believes,” “projects,” highlighting in this way the extreme power it bears on humans in taking control of all human interpretations (Nietzsche, 21). Becoming the single arbiter of the self’s relation to itself, others, the world, its occurrences, and all objects within it, reason appears to grow almost too powerful, undermining human authority and autonomy.

When our understanding of the world is solely reliant on rational concepts, our perspective becomes severely restricted, leading to a loss of self-determination. By being a powerful agent in assessing our surroundings, reason creates the world as it appears to us, making sense of the chaos we were thrown into. Underlying and permeating every aspect of our interacting with the world, the faculty of reason expresses itself through the creation of the illusory self, which subsequently posits itself as the centre of every perception and experience, dispossessing humans of their autonomy.

Moving Beyond a Cartesian Framework and Reclaiming Agency

As reason constructs the world by recognition, classifying it as “deeds and doers,” objects and subjects, or causes and consequences, it creates an illusory perfect reality (Nietzsche, 21). Its authority is almost invasive. Reason transcends borders “everywhere,” and “on to all things;” it interferes with the outside world and intrudes, but not always with consent (Nietzsche, 21). Although reason’s intention is to find explicative causes to interpret and explain humans’ lives, its *hubris* causes it to be omnipresent, taking away humans’ agency. The perverted aspect of reason restricts and destroys passions and pulsions, endangering what Nietzsche believes to be necessary to the accomplishment of the will to power (*Wille zur Macht*). The will to power is what provides meaning and direction to human existence, restlessly overcoming one’s obstacles and transcending the self to achieve one’s goals. Language further reinforces this illusion of a stable world, as it can act as a confining structure, enforcing a subject-predicate framework, thereby perpetuating the misconception that reality is arranged and that absolute truths exist. Language and reason, which humans created as means for emancipation and upliftment, result in a deeper restriction than previously experienced.

It would be foolish to grant absolute free will to human beings if the concept of ‘ego’ is fundamentally constructed by reason itself. Indeed, Johann Gottlieb Fichte, associated with German idealism, argued that reason is a self-pos-

iting, active force, necessary to construct knowledge and moral freedom. Rejecting previous dogmatic truths that establish the ego (*Ich*) as the passive source of certainty, Fichte wrote in his *Wissenschaftslehre* that the ‘I’ is the origin of all experiences driving moral and rational actions. However, grounding all experiences in the ego seems to create a paradox: while it serves as the source of human determination, it simultaneously imposes limits on the possibility of acting freely. According to Nietzsche, being appears to be an alien concept, void of inherent meaning. Reason, as an artist, imbues colours onto reality, assigning it value and significance. From the concept of ‘ego,’ which emerges as a perceived “substance,” arises the fixed and independent framework of perception that culminates in the concept of ‘being’ (Nietzsche, 21). The ego projects itself outward, orienting itself toward a future in which it assumes its existence. In doing so, it tangibly manifests, and reshapes, everything it encounters. Henceforth, the ego becomes the cradle of interpretation, that through which is born the force of human perception.

Ego, as an artificial ‘conception,’ projects its own images and self-interpretations onto reality (Nietzsche, 21). Consequently, what is external or not previously contained within the ego appears alien and must be subordinated to it. In this sense, Nietzsche claims that reason does not liberate us from animality as suggested by German idealists, but rather confines us within it. If ego is “everywhere thought in, [and] foisted on,” then humans are unable to escape its grasp (Nietzsche, 21). Almost trapped, the self—through reason—imposes itself upon the world, objectifying what surrounds it. Whether in thought or in action, the ego prevails. In this way, the absolute and insatiable necessity to search for causes and to rely on reason as an absolute guiding force is, according to Nietzsche, a constraint rather than a means of liberation.

Nietzsche asserts that humans are shaped by multiple elements, notably by their culture and environment, rather than existing independently. Despite this, it seems that humans often perceive themselves as entirely responsible for their actions: Nietzsche rejects this predominant world-view, asserting that the concept of the self-directing ‘I’ is an illusion. Against his predecessors, Nietzsche argues against the notion of the self as an autonomous force, rather suggesting that the self did not arise from the vacuum, but was, in fact, the result of determinant exogenous factors. Therefore, thinking of ‘I’ as an indeterminate concept is absurd, as it is in perpetual construction. Furthermore, humans’ ongoing en-

slavement to reason started when they chose to consider the will as a faculty. This is, in Nietzsche's terms, "the great fateful error" of humanity, leading us astray (21). He suggests the will should be understood not as being inherent, but rather as a product of cultural influences, including desires, environments, social classes, religion, and language. Hanging on to the conception that 'I' is entirely free from influences merely puts a veil on the true nature of human existence and is unhealthy, making humans sick with weakness of will. This view pushes humans into thinking that their thoughts and actions are free from influences, whereas external forces constantly forge them, building their behaviours and speech. If humans do not use their sense of critical thinking and passively exist in the world without challenging it, they will inevitably live in subjugation of their own illusions, failing to protest when their freedoms are being diminished.

Conclusion

To conclude, Nietzsche criticizes traditional notions of reason and rationality, perceiving them as obstacles to human flourishing. Indeed, language, this arbitrary tool allowing communication between individuals, is repressive as it fails to translate the deepest meanings of reality. Arguing that reason, as traditionally understood, has often been used to justify and reinforce oppressive ideologies and belief systems, Nietzsche attacks German idealism. He advances that their logical systems, based on the manifestation of reason, show the interconnection between language and reason in trying to obscure the chaotic nature of the world. Nietzsche suggests that reason, in its excessive categorization, leads to the demise of humanity and its freedom. To aim for a new and healthy age, Nietzsche offers to embrace the will and assert human nature without limitations through creative means in the intellectual and spiritual realms. Beyond the limitations of rationality, Nietzsche emphasizes the importance of transcending social expectations to embrace natural impulses, for they represent human nature, shaping our experiences within the world.

Works Cited

- Fichte, Johann Gottlieb. *J.G. Fichte: Foundation of the Entire Wissenschaftslehre and Related Writings (1794–95)*. Edited and translated by Daniel Breazeale, Oxford University Press, 2022.
- Hegel, G. W. F. “Phenomenology of Spirit.” *G.W.F. Hegel: The Oxford University Press Translations*. Translated by A. V. Miller, edited by J. N. Findlay, electronic ed., InteLex Corporation, 2000.
- Kant, Immanuel. *Critique of Pure Reason*. Translated by Norman Kemp Smith, St. Martin’s Press, 1965. Originally published 1787.
- Nietzsche, Friedrich Wilhelm. *Twilight of the Idols: Or, How to Philosophize with the Hammer; The Anti-Christ; Notes to Zarathustra and Eternal Recurrence*. Gordon Press, 1974.



On the Possibility of Nonviolent Decolonization: Fanon, in light of Gandhi

VALENTIN PELOUZET

In this essay, I will attempt to answer the question “What is the role of violence in the decolonial struggle?” My analysis will be centred around psychiatrist and philosopher Frantz Fanon’s work, *The Wretched of the Earth*, which was published in 1961, amid the Franco-Algerian war. I will first explore and analyze the role of violence in decolonization for Fanon, drawing particular attention to why he finds ‘absolute violence’ to be a necessary means for decolonization. Later, I will evaluate this argument, not from a moral but from a strategic perspective. I will question whether Fanonian violence is the only, or even the preferable means for decolonization. Overall, I will argue that, from a practical perspective, violence should not be the preferred means of struggle for decolonization. Rather, I will argue in favour of the use of non-violence, notably inspired by Gandhi’s theory of *satyagraha*. More precisely, I will show that Fanon’s criteria for successful decolonization can be met more efficiently through nonviolence.

Keywords: violence, decolonization, Fanon, psychology, colonization, nonviolence, Gandhi, satyagraha

Introduction

By 1961, the publication year of *Les Damnés de la Terre* (*The Wretched of the Earth*), Frantz Fanon was well aware of the fragility of the decolonial enterprise. Written from Algeria, amid a war led by the FLN (Front de Libération Nationale) and its armed branch, the ALN (Armée de Libération Nationale) with the French (Zeilig, 11), the book can be interpreted as a warning to Algerians, and later decolonial movements.

A warning, first, on the mistake of choosing the route of elite-driven diplomacy for independence. Alongside Jean-Paul Sartre, who signs the book's preface, Fanon strongly doubts the colonial bourgeoisie's capacity and willpower to create a postcolonial nation, economically, culturally, and psychologically free from the colonizer. For the two authors, these negotiations lead to a "farce of national independence" ("farce de l'indépendance nationale"; Fanon, *The Wretched*, 67)¹ which has nothing to do with decolonization, which requires a full rupture of the subjects of colonialism from the intertwined colonial infrastructure and super-structure.

Rather, Fanon advocates for "absolute violence," ("violence absolue", my trans.) (a decolonial war) as the necessary means for the establishment of the postcolonial nation. Fanon famously opens *The Wretched* by writing that "decolonization is always a violent phenomenon" ("la décolonisation est toujours un phénomène violent"; 35) and dedicates its first part to the defence of decolonial violence. Whilst Fanon widely discusses symbolic and cultural violence in his texts, here I will focus on decolonial violence as physical confrontation, ranging from colonial armed oppression to guerilla tactics and terrorism (Seshadri-Crooks, 84). I will justify this decision in Part I.

This defence of violence is certainly political, in that Fanon is deeply worried about the colonizer's sustained domination of the colonized masses, through the intermediary of the colonial bourgeoisie; he has witnessed this continuation of colonial dynamics in Senegal and Ghana (Zeilig, 13). Violence, for Fanon, is the means through which the masses are involved in decolonization (Fanon, *The Wretched*, 123-4), and through which they fully break from the colonial metropole (41).

However, as I will show in Part I, this defence is also heavily psychological. Here, by 'psychological,' I mean all the effects violence has on colonized

1. All translations for this text were made by me.

masses' psyches, including their personal and interpersonal mental saneness, as well as their relationship with and view of the nation. Fanon believes that decolonial violence is essential for the masses' physical liberation, their freedom from alienation—which I will define later—and for the development of their national conscience. My focus in this paper will be the assessment of these three psychological justifications of violence.

As a note, as most interpretations agree on, Fanon holds more nuanced views on violence than commonly portrayed. In Part II of *The Wretched of the Earth*, he details the weaknesses of “spontaneity” (“spontanéité”), the spontaneous and violent mobilization of rural masses in the struggle (105). Subsequently, in Part V, Fanon details and exemplifies the deep consequences of violence on the psyche—a discourse doubtlessly influenced by his training in psychoanalytic psychiatry. I will expand on both of these arguments later, but I want to insist on the following point: the use of the term “wretched” acts as a warning of the dangers of the necessary, inescapable, yet uncompromising violence decolonization requires.

This paper aims to further inquire into the psychological dangers of decolonial violence. Ultimately, I will argue that violence should not be the preferred means of struggle for decolonization. For this purpose, I will try to go beyond the Fanonian dichotomy between elite negotiations, leading to “phony independence” as Sartre calls it (“une fausse indépendance”; 20), and “absolute violence,” leading to decolonization. In other words, I will not try to refute Fanon's position on elite-driven independence here; rather, I will try to show that another route to decolonization is possible: nonviolence.

While nonviolent social theorists' works mostly responded to theorists of violence, such as Fanon, they surprisingly have faced very few critical responses from the latter. As Brian Martin writes, “nonviolence is most commonly ignored by its critics” (Martin, 83). Indeed, nonviolent social movement theory is often interpreted as, and limited to a “humble appeal to conscience” (Deming, 5). The emergence of this view can at least partly be explained by the fact that the fundamental immorality of violence is the key, but not sole, argument for several prominent nonviolent theorists. One may name among them Gene Sharp and Theodore Paullin, as well as activists including A.J. Muste and Martin Luther King, Jr. (R. Howard, 9-10).

In this essay, while my approach to the question of violence in decoloni-

zation will mainly be normative, my argument is *not* primarily moral. The choice to focus on the deeply political context of decolonization holds me back; as Yves Winter writes, moralization is deeply depoliticizing, as it “typically brackets considerations of causes, dynamics, and implications” of violence (Winter, 5). I will thus not allow myself any moral judgment onto those who, in their struggle for self-determination, chose to fight violently. My essay, rather, will act as a recommendation for the sake of colonized people’s personal and national sanity and liberation.

For this purpose, I will base my understanding of nonviolence on Mahatma Gandhi’s *satyagraha*, or nonviolent action (Mantena, 455), which he developed through the struggle for India’s decolonization (N. Howard, 101). This choice to use the Gandhian paradigm serves two purposes in the context of the point enunciated above. First, *satyagraha* was developed through a decolonial struggle, which offers a helpful comparison with Fanon’s theory. Second, its main considerations are arguably not moral, but political and practical. As Karuna Mantena writes, *satyagraha* should be interpreted as a “plausible practical orientation in politics and not purely a moral proposition, ethical stance, or standard of judgment” (Mantena, 458).

This essay will go as follows: I will first analyze the psychological roots of decolonial violence in Fanon and its justifications, focusing on the three arguments enunciated above. I will subsequently offer a critique of these arguments, drawing on Fanon’s works and his critics. Finally, with the use of the Gandhian literature of *satyagraha*, I will show the crucial qualities and corrections which nonviolent action offers. Overall, I will argue that violence should not be the preferred means of struggle for decolonization. More precisely, I will show that the psychological objectives Fanon seeks to accomplish at with violence are better attained through nonviolence.

Part I - The “Grandeur” of Violence in Fanon

Fanon’s depiction of violence can be interpreted both descriptively and prescriptively. Indeed, an important component of his argument is that violence is already entrenched within the subjects of colonization since the colonial *status quo* is inherently and profoundly violent. In this context, decolonial violence acts as the mirror of colonial violence. Fanon writes of this, “The violence of the colonial regime and the counter-violence of the native balance each other and

respond to each other in an extraordinary reciprocal homogeneity” (“La violence du régime colonial et la contre-violence du colonisé s’équilibrent et se répondent dans une homogénéité réciproque extraordinaire”; Fanon, *The Wretched*, 85). To understand his argument, it is thus essential to analyze the latter.

This violence is at first cultural: the colonial world, Fanon explains, is “a Manichean world” (“un monde manichéiste”; *The Wretched*, 44). In the colonial setup, there is an essential, immediate, and strict distinction made through race, between the White and the Indigenous. The latter is deemed amoral and “insensible to ethics,” (“imperméable à l’éthique”; 44) and becomes, through this process, closer to an animal than to a man. By extension, Indigenous cultures come to be considered as inherently inferior, and are methodically destroyed by the colonizer (44-5).

A pernicious effect of this essentializing inferiorization is that it is interiorized by the colonized. This interiorization is described by Fanon in his other major work, *Peau Noire, Masques Blancs* (*Black Skin, White Masks*). He explains that the colonized develop an “inferiority complex” (“complexe d’infériorité”; *Black Skin*, 8)² from their constant inferiorization. They come to internalize and accept the European’s superiority, going so far as to envy them, this new reality being one in which “the Black wants to be White” (“le Noir veut être le Blanc”; 7). Colonial violence is thus also symbolic—in the Bourdieusian sense—for it “integrates into cognitive structures and operates with the complicity of its victims” (Landry, 85). Similarly to Pierre Bourdieu’s notion of class-based violence, colonial violence expresses itself (to some extent, not at all entirely) through social norms of racial hierarchy which the colonized come to integrate and apply themselves.

To this extent, the colonized alienate (from “*alius*”, the other) themselves, or become foreign to themselves, in that their existence becomes relative to the colonizer. As Fanon writes, “It is the colonizer who *makes* the colonized” (“c’est le colon qui *fait* [...] le colonisé”; *The Wretched*, 40). Through this process, the colonized become an object of colonization and thus lose their agency and freedom. This alienation happens on two levels: cultural and material. On the cultural level, “[they] see European culture as a means of freeing [themselves] from their race”; whereas, on the material level, “it is as a victim of a regime based on the exploitation of one race by another” (“c’est en tant que victime d’un

2. All translations of this text were made by me

régime basé sur l'exploitation d'une certaine race par une autre") that they alienate themselves (Fanon, *Black Skin*, 181).

Indeed, while the superstructure—the colonial culture—is in itself extremely violent, it ultimately serves as a validation of a colonial dichotomy, the cultural setup acting as justification for economic inequality. As Fanon indicates, in colonialism, “the economic infrastructure is also a superstructure” (“l’infrastructure économique est également une superstructure”; *The Wretched*, 43). This inequality is materially violent in itself, but violence becomes purely physical (notably, but not exclusively) when the essential inequality is contested. In Fanon’s words, “barracks and police stations” (“les casernes et les postes de police”; 41) guarantee the compartmentalization of the colonial world. While the inequality in itself is justified by the overall inferiority of the colonized, the use of force is legitimized by their animality, or incapacity to understand any form of communication other than “the language of force” (“le langage de la force”; 81).

Since Fanon’s understanding of colonial violence is physical and material on one hand, cultural and symbolic on the other, as well as both infrastructural and superstructural, one may wonder why I chose to treat Fanon’s decolonial violence only through the case of physical confrontation. This is even more puzzling in that, in the opening of *The Wretched of the Earth*, decolonization is presented as also being a culturally violent phenomenon. Fanon uses “the new names of sports clubs” (“[les] appellations nouvelles de clubs sportifs”; 39) or “the human composition of cocktail-parties” (“[la] composition humaine des cocktails-parties”; 39) as examples to show that the population replacement decolonization implies is violence in itself. In light of this, some readers of Fanon, including Barbara Deming, have interpreted Fanon as possibly using the term “violence” to mean “radical and uncompromising action” (Deming, 2).

Nevertheless, the bulk of Fanon’s discussion of decolonial violence, in the text, concerns physical violence. Fanon repeatedly associates violence with the call to “take arms” (“prendre les armes”; *The Wretched*, 72, 123, 127): I will therefore dedicate an important portion of Part II to the discussion of military tactics. The very beginning of *The Wretched* does not seem representative of Fanon’s discussion of decolonial violence; the cultural violence of decolonization seems rather to be only a corollary of the colonial war. From my understanding, Fanon wrote this incipit to immediately show to the reader that decolonization, defined as “the replacement of a species of man with another” (“le remplacement

d'une espèce d'hommes par une autre"; 40), is inherently violent, at least culturally, before segueing into a more precise defence of physical decolonial violence. I will thus focus my analysis on an understanding of Fanonian violence as physical violence.

I find three main justifications for violence in *The Wretched*. The colonized subject's violence is first, as Michael Sonneleitner writes, a "physiological response to internal tensions that, if not released externally, would result in a high occurrence of self-destructive forms of 'mental pathology'" (290). Armed struggle allows colonized individuals, at the personal level, to relieve their aggression, accumulated through years of oppression and suffering. As mentioned before, violence, or at least the potential for violence, is already present in the colonial status quo; the colonized masses only channel it through different means, such as "dance and possession" ("la danse et la possession"; Fanon, *The Wretched*, 57), or "fratricide struggles" ("les luttes fratricides"; 59), i.e. homicides or clan wars.

This argument is undoubtedly linked to Fanon's experience as a psychiatrist. In Part V of *The Wretched*, he relates multiple cases in which homicidal tendencies resulted from colonial violence. Case n°2 of series A, for instance, describes an Algerian man, foreign to politics, who, following a "collective liquidation" ("liquidation collective") of Algerians by French soldiers, became subject to some "undifferentiated homicide impulses" ("pulsions homicides indifférenciées"; Fanon, *The Wretched*, 248-250). The term "undifferentiated" is key here; this man, whom Fanon describes as non-politicized and unconcerned by the war, does not direct his violence in a politically productive direction. The colonized's violence is an inevitable consequence of the colonizer's violence; the nature or identity of its target, however, is not.

In this context, the belligerence of decolonization comes to be a reorientation of this aggressiveness toward a legitimate cause. The violence which "r[an] idle" ("tourn[ait] à vide"; Fanon, *The Wretched*, 59) now "confronts the only forces that challenged its existence: those of colonialism" ("affronte les seules forces qui lui contestaient son être: celles du colonialisme"; 58). Through this process, the colonized no longer require the outlets through which they previously channelled their violence; Fanon observes "a peculiar disaffection for these practices" ("une désaffection singulière pour ces pratiques" ; 58) in wartime. This disaffection, for Fanon, is the proof of the violence transfer decolonial violence provokes.

Relatedly, violence is a means for the colonized to realize their fun-

damental equality with the colonizer. For Fanon, acts of violence “prove to the oppressed that the colonizer is as mortal as they are and that they are as powerful” (Sonnleitner, 290). In this sense, it is a negation of the inferiority complex which characterizes the colonized during the colonial period. Violence against the author of this essentialization is thus a symbolic and performative means to “question the colonial status quo” (Fanon, *The Wretched*, 40).

This realization of equality is simultaneous with a realization of existential liberty. As Parris indicates, Fanon is deeply influenced by existentialism (5); he believes that men are fundamentally free and able to create their own purpose. However, for Fanon, the colonized’s understanding of this freedom is frustrated by their alienation. Through violence, they realize their power to act on the oppressive *status quo*, correcting this “existential deviation” (“déviation existentielle”; Fanon, *Black Skin*, 11). More importantly, through this process, the colonized, “illuminated by violence” (“illuminé[s] par la violence”; Fanon, *The Wretched*, 91), become “new men” (“des hommes nouveaux”; 40).

Fanon’s existentialism, following the thought of Simone de Beauvoir and her influence on Jean-Paul Sartre, is not to be understood as a defence of individualism; on the contrary, that decolonisation “fundamentally modifies the being” (“modifie fondamentalement l’être”; Fanon, *The Wretched*, 49) into “new men” (“hommes nouveaux”; 49) ultimately serves the collective creation of the nation.

Fanon’s final argument in favour of decolonial violence is that it is necessary to create the nation: the decolonial war redirects “fratricide” and other forms of internal violence onto the colonized; it acts as a means of unification of the masses (Fanon, *The Wretched*, 86-9). As the war begins, and especially when the colonial power retaliates—which marks the “point of no return” (“point de non-retour”) at which a ‘cycle of hatred’ inevitably starts (86)—the unification of the colonized against their oppressor becomes blindly antagonistic. Through this process, ironically, the colonized appropriates the colonizer’s Manichaeism: “The theory of the ‘Indigenous as absolute evil’ is met with the theory of the ‘colonizer as absolute evil’” (“À la théorie de « l’indigène mal absolu » répond la théorie du « colon mal absolu »”; 89). The colonized continue to define themselves dichotomically, in opposition to the colonizer, but collectively, with pride, and for liberation. The “national conscience” of the masses, incapable of emerging under bourgeois leadership, emerges from this Manichean struggle (Sonnleitner, 296-7).

The Manichean War of Liberation acts for Fanon as the founding my-

thology of the postcolonial national culture, which “manifests itself on the battlefield” (“se manifeste dans le combat”; Fanon, *The Wretched*, 212). At this point, it is important to remember that the colonial enterprise actively crushes Indigenous cultures. The postcolonial nation thus requires a narrative onto which its newly acquired national conscience can lie. As Seshadri-Crooks writes, violence thus “gets legitimated primarily through its elevation or incorporation into the narrative of national liberation” (85). This national narrative becomes a “meta-narrative” (88) to the extent that it is built performatively by nationally conscious masses.

Part II - “Faiblesses de la spontanéité”, and other critiques

As Hannah Arendt writes, Fanon “is much more doubtful about violence than his admirers” (14). As was mentioned in the introduction, being a direct witness of war (and an actor in it, during WWII), as well as a psychiatrist, Fanon was well aware of the psychological dangers and damages caused by war and the kind of violence it entails.

In Part II of *The Wretched*, he details the dangers associated with the spontaneity of decolonial violence (105). As explained earlier, Fanonian violence is not only strategic; it is already a fact of the *status quo* and is redirected towards an ideally unified antagonist: the colonizer. In this context, violence can become uncontrollable. The call for violence tends to come early in the masses’ political awakening (Fanon, *The Wretched*, 124): they are “impatient”, “reject any nonviolent method” (“impatient[es]”, “rejette[nt] toute méthode qui ne soit pas violente”; 72), and answer to their “bloodthirsty instincts” (“instincts sanguinaires”; 71). I will here show how this absolute unleashing of violence can be detrimental to the ends Fanon pursues, especially regarding the aftermath of the decolonial war.

The first disadvantage of decolonial violence concerns mental disorders, contrasted with the physical liberation they allow. The last part of *The Wretched of the Earth* is dedicated to the mental disorders linked to decolonial war. Whilst the section shows in great detail how being a victim of violence in decolonial war can create psychological disorders, it is evasive about how the experience of violence as an executioner, such as the “post-traumatic disorder” (“trouble réactionnel” Fanon, *The Wretched*, 242), can create psychological distress; he only mentions that the assumed “benignity” (“bénignité”) of these “disorders” (“troubles”) should be “slightly relaxed” (“légèrement assouplie”; 242).

Fanon's lack of consideration for the psychological state of the colonized highlights his reluctance to incorporate this key weakness into his analysis of violence. To expand on the question of post-traumatic stress, however, it is interesting to look at the case he associates with this 'disorder', which he only mentions in a footnote. He recalls the example of a FLN militant, who, after placing a bomb in a café, started suffering from what Fanon calls "a kind of dizziness" ("une sorte de vertige"): the militant becomes terrified that those he murdered could have very well been ex-inhabitants of French Algeria he had befriended (243).

Picked up by Deming (1), this illustration highlights the fact that, past the frenzy of vengeance and physical liberation, and more generally the war, violence has a heritage. In other words, "our actions never cease to follow us" ("nos actes ne cessent jamais de nous poursuivre"; Fanon, *The Wretched*, 243). However, Fanon does not frame the issue of our acts' 'haunting' in the wider context of the postcolonial nation's heritage (243). The collective war trauma is indeed exclusively limited to the French's atrocious persecutions—"France's human heritage in Algeria" ("l'héritage humain de la France en Algérie"; 241). He makes no mention that this heritage is determined by the use of violence by Algerians as well.

There is indeed a structural, collective legacy of decolonial violence. Fanon analyses decolonial violence as a way to liberate oneself from the inferiority complex one developed during the colonial period—violence as a tool to free the colonized from their essentialization. Nevertheless, the fact that part of this essentialization takes the form of animalization (Fanon, *The Wretched*, 45) undermines this argument; the use of violence acts as a confirmation of prejudice, as through its use the colonized confirms that "he only understands the language of force" ("il ne compren[d] que le langage de la force"; 81).

Fanon attempts to solve this tension by arguing that the decolonial war shows the colonizer that they, themselves, only understand violence (Fanon, *The Wretched*, 81). Problematically, this response shows that the use of violence acts more as a devalorization of the colonizer rather than as a valorization of the colonized. This confusion departs from the fact that Fanon generalizes violence as an "indissoluble" necessity in rebellion (Castelli, 335), and altogether as a necessary means to claim power. If this assumption is true, the use of violence shows the colonized that they are as capable of asserting power as the colonizer is.

As Hannah Arendt details, however, the idea of violence as the necessary

source of power, following the maxim “the law cannot be enforced without the sword” (Wiley, 332), is itself traditionally Western—and colonizing. By recuperating the methods of the European colonizer, the colonized, even if victorious, never fully emancipate themselves from the colonizer. The Western superstructure is in this way not eradicated by the struggle, as Fanon suggests (*The Wretched*, 49), but rather reproduced and validated. As Green concludes, “To resist violently would be to yield to the conqueror’s ideology” (195).

More generally, Fanon’s view of violence as inevitable, rather than ideologically motivated, seems contradictory with his ideal of the existentially free ‘new men.’ By considering violence, against the colonizer as well as in other forms, as a fatality, he himself essentializes the colonized. The latter, through violence, only confirms that they are slaves to their primal drives, condemned to remain alienated, not by the colonizer but by their impulses. As Kebede writes, “[Fanon’s] therapy of violence needs to explain how capitulation to an impulse amounts to a victory over that impulse” (558).

Advocates of Fanon could answer that the colonized is not existentially free, and is unable to create their own purpose until they have liberated the violent impulses they hold against the colonizer. However, as Fanon notes himself, violence does not disappear after the decolonial war, for “the young independent nation evolves during its early years in an atmosphere of a battlefield” (Fanon, *The Wretched*, 93). And one who upholds Fanon’s theory must recognize that Fanon himself is aware of the risk of violence’s uncontrollability, going so far as to state that “hatred cannot constitute a program” (“la haine ne saurait constituer un programme”; 134), suggesting that violence is necessary, but needs to be framed around political aims.

Fanon perceives this violence as a creative force, a “creative madness” (“folie créatrice”; Fanon, *The Wretched*, 93), and not as a danger. On the surface Fanon agrees with Gandhi who, like him, believed that the means of struggle are “ends-creative,” in that they actively shape the ways in which ends come about, particularly regarding the institution of a national conscience (Gandhi, 468). Nonetheless, for Gandhi, violence does not create the conditions for national unity to arise. On the contrary, it feeds itself in a “cycle of violence” in that it “necessarily puts into motion chains of animus and dissension that ultimately result in instability” (466).

As Arendt points out, violence creates the risk that “the means over-

whelm the end” (80): violence may go beyond its instrumental role, and become an end in itself in the postcolonial society. This is especially concerning, in the context of national conscience, regarding Fanon’s disdain for “fratricide struggles” (“luttres fratricides”; Fanon, *The Wretched*, 59). As Arendt notes, the national fraternity which stems from the decolonial struggle is fragile, for it is mostly bound by the existence of adversity which equalizes all before death (Castelli, 335). This formal equality disappears with the war, and the national conscience may not resist this vanishing. The “brotherhood on the battlefield” (335) therefore seems to be by nature transitory.

In this sense, there is a strong likelihood that, after the war, this violence will express itself among the ex-colonized, within the postcolonial nation. As Judith Butler concludes in their analysis of Sartre’s preface, “there is a question of whether violence as a pure instrument can remain such, or whether it comes to define, haunt, and afflict the polity that instates itself through violent means” (225).

Part III - Gandhi’s satyagraha: The possibility of nonviolent decolonization

I have detailed the structural issues linked with violence in decolonization. They, however, are not enough to argue against its use. In fact, an obvious response could be that violence is the sole alternative to elite negotiations in decolonization. I will here argue, through Gandhi’s notion of *satyagraha* and following Butler’s critique of Sartre’s preface (227), that this assumption of a binary choice is wrong. More precisely, I show that nonviolence is a credible alternative to absolute violence in the decolonial context.

Nonviolence is often disregarded in radical theory. Fanon himself only mentions it once in *The Wretched* and limits it to a bourgeois tactic stemming from colonial elites (61). This confusion, Deming notes, stems from the assumption that “a [radical] challenge can only be offered violently” (4). Radical theorists, including Fanon, often assume that the violence of *means* should be proportional to the radicality of their *ends*. As has been shown previously, however, radical means can undermine the radical ends they serve in the context of decolonization, especially post-war.

Gandhi, deeply concerned with the overall relationship between means and ends (Mantena, 496), aims at similar radical ends as Fanon. As Niel Howard suggests, both Gandhi and Fanon agreed on the “ultimate goal” of decolonization: the “freedom of the nation” (96). Both believed that “personal [and] national

reconstruction” are deeply linked to this freedom, and that, in this context, there is a “crucial interplay between the individual and the collective” (95-6) in the decolonial struggle. Finally, both insisted on the “absolutely integral” (98) role of national culture and values, both as a colonial tool of domination and as a means for emancipation. The psychological dimension of the colonized subject, though framed with different vocabularies, is thus key to both Fanon and Gandhi’s theory of decolonization.

Gandhi and Fanon’s point of departure is the same: the reality of colonial violence (N. Howard, 95). This violence is obviously physical and material, but also cultural and symbolic; just as Fanon’s time in the French metropole, Gandhi’s experience in London gave him a “sense of exclusion and inferiority” (95). Both also aligned regarding the potential for violence within the colonized. For Gandhi, violence, or himsa, is “a dynamic feature of political contestation” in that it “initiates definite dynamics of resentment, retrenchment, and retaliation” (Mantena, 464). The colonizer’s violence thus feeds into a “cycle of violence” (466), which drives the potential for the violence of the colonized.

The authors diverge in their prescriptions regarding radical violence as a means. Rather than the liberation of this violence onto the colonizer, Gandhi advocates for the colonized’s self-suffering. He defines *satyagraha* as a discipline of conscious suffering, in which the *satyagrahi* (the nonviolent actor) endures the colonizer’s violence. They are not passive; their contestation is framed as a sacrifice, a full bearing of “the burden of acting” (Mantena, 470). They demonstrate their convictions by fully and publicly embracing the risk of the severe consequences they face.

From this understanding, *satyagraha* avoids all the issues we have seen in the second part of this essay. Self-suffering during the struggle avoids the post-traumatic suffering of inflicting pain. The “cycle of violence” (Mantena, 466) is stopped by the *satyagrahi* and thus does not animate post-war resentment. The nonviolent actor may, in this way, truly liberate themselves, both from their inferiority complex and from their impulses.

This may be understood in two ways. On one hand, the decolonial practice of *satyagraha* disalienates the colonized by giving them mastery over their animality, or “passions” in Gandhi’s work (Mantena, 455). The *satyagrahi* overcomes the anger and fear which would lead them to violence, respectively as a reaction and as a defence. They prove the colonizer wrong by refusing to

react as they are expected to. They, in this way, make a truly *existential* statement, use Fanonian paradigms, and demonstrate their radical freedom both from the colonizer's coercion and from the cyclical logic of violence which characterizes politics.

Relatedly, the *satyagrahi* is alleviated from their inferiority complex through their refusal to adopt the colonizer's values. The logic of violence and counter-violence is deeply colonial: through nonviolence, the colonized "refuse [the colonizer's] world, and decontaminate their soul by the affirmation of positive human values" (Kebede, 559). In the process, they prove the colonizer's world's logic wrong and show that the lack of consent to oppression can be shown actively and effectively without violence. It is in this demonstration that nonviolence mainly finds its effectiveness.

A tempting Fanonian response to Gandhi would be that *satyagraha* would rapidly bend in front of the colonial power. The topic of this paper, again, is not the material capacity of different struggle tactics in the decolonial context. Related to our psychological focus, it is important to note the converse (opposed to coercive) power of *satyagraha*, i.e. its power to convert. Whilst this is not its sole lever of action, nonviolence stems from the belief that nonviolent struggle can "raise the level of consciousness" of the antagonist (Deming, 7). In respecting the adversary's (the colonizers) rights, but not their policies, "the focus of attention becomes those actions, those policies, and their true nature" (7). The colonizer, not personally threatened but observing the violence with which these nonviolent actions are met, has no choice but to become conscious of the colonial system's violence rather than the colonized's.

Another Fanonian response would be to question the possibility of the masses' implication in the decolonial project. Violence, for Fanon, is a useful device in that it is already present within each colonized person—it can thus be summoned and redirected according to the needs of the movement. In Gandhian theory, concerns of popular implication are equally important. However, the masses' mobilization, and to an extent, the awakening of the national conscience, happens through a "regaining of individual self-awareness and thereby an Indigenous cultural identity" (N. Howard, 98). The masses' awakening also comes at the "intimate scale" (Mantena, 475) but through the reminiscence of their pre-colonial identity.

Fanon is very critical of the call to Indigenous cultures: "To cling to

tradition or to revive abandoned traditions,” (“Vouloir coller à la tradition ou réactualiser les traditions délaissées,”) he writes, “is not only to go against history but also against one’s people” (“c’est non seulement aller contre l’histoire mais contre son peuple”; *The Wretched*, 213). He believes that the postcolonial culture necessarily arises through the struggle and that the masses do not relate to what he claimed to be “exotic” (“exotique”; 210) calls to indigeneity. If for Gandhi, the cultural struggle is the basis through which the population becomes materially invested in the decolonial enterprise, for Fanon the struggle rather begins from the physical or material (the war), and through it becomes cultural (through the construction of the national conscience).

The masses’ relation to precolonial culture may be dependent on the extent of their destruction by the colonizer. However, as Partha Chatterjee showed through anthropological and historical research, anticolonial popular nationalism can actually begin with cultural nationalism. It comes, as suggested by Gandhi, from the appropriation of “[the] domain of the spiritual”, or the “inner domain”, before the struggle in the “outside” (Chatterjee, 216-7). Chatterjee takes the example of the “Bengali literary renaissance”, or the revival of the use of the traditional Bengali language in the region as a countercultural operation against Britain. Sovereignty here was first and foremost claimed through the inner domain of language, and the spiritual”, or the “inner domain”, before the struggle in the “outside” (Chatterjee, 216-7). Chatterjee takes the example of the “Bengali literary renaissance”, or the revival of the use of the traditional Bengali language in the region as a countercultural operation against Britain. Sovereignty, here, was first and foremost claimed through the inner domain of language, an essential part of traditional culture (218-9). Indigenous cultural richness, therefore, need not be definitively abandoned in favour of a mythification of violence; on the contrary, it can be used as the fundamental rejection of the colonizer’s inferiorization and domination overall.

Conclusion

Through this paper, I have shown why nonviolent tactics, mainly understood through Gandhi’s theory of *satyagraha* and its legacy, should be privileged in decolonial struggles. My focus here was the psychological; I showed the psychological objectives Fanon aims at with violence, namely the physical liberation, freedom from alienation, and destruction of the inferiority complex characteristic of the colonized, as well as the development of a national conscience within the

colonized, are better attained through nonviolence. My argument was both destructive, in contradicting the viability of Fanonian violence for the goals he aims at (Part II), and constructive, in showing how nonviolence better serves these goals (Part III).

I only hinted at the strategic and practical implications of violence and nonviolence—I have not fully discussed how they act and compare to defeat the colonizer materially. This discussion would deserve a full other paper. The colonized’s psychology remains a key dimension of both Fanon and Gandhi’s work. Whilst they ultimately disagree, this concern shows both author-activists’ profound implication and empathy towards victims of colonization.

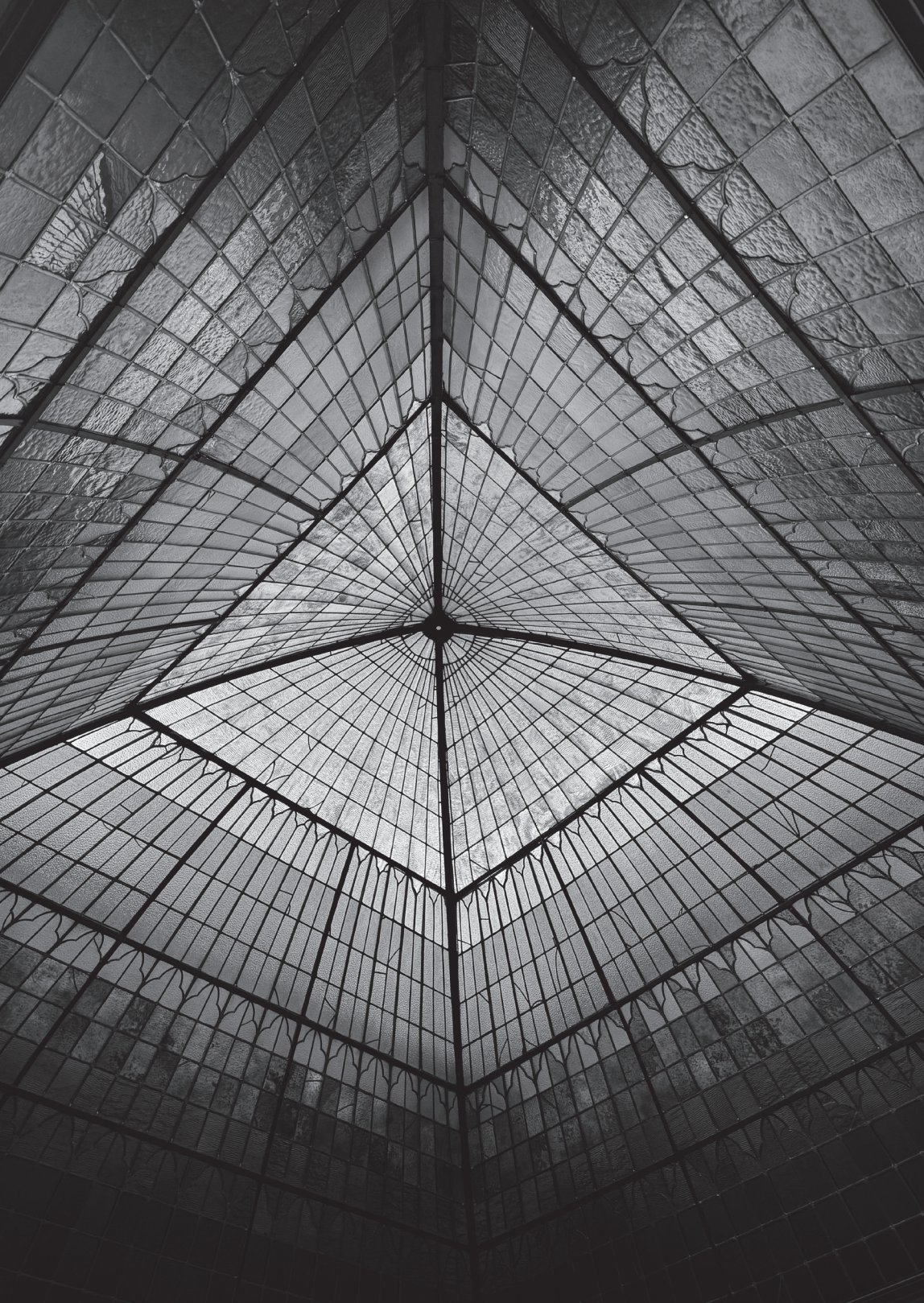
Works Cited

- Arendt, Hannah. *On Violence*. Harvest/HBJ, 1969.
- Butler, Judith. "Violence, Nonviolence: Sartre on Fanon." *Race after Sartre: Antiracism, Africana Existentialism, Postcolonialism*, edited by Jonathan Judaken, SUNY Press, 2008, pp. 211–231. <https://doi.org/10.1515/9780791477854-012>
- Castelli, Alessandro. "Liberation through Violence in Fanon's *The Wretched of the Earth*: Historical and Contemporary Criticisms." *Peace & Change*, vol. 47, no. 4, 2022, pp. 325–340. <https://doi.org/10.1111/pech.12554>
- Chatterjee, Partha. "Whose Imagined Community?" *Nations and Nationalism*, Edinburgh University Press, 2005, pp. 237–247. <https://doi.org/10.1515/9781474472777-022>
- Deming, Barbara. "On Revolution and Equilibrium." *Liberation*, vol. 12, no. 3, 1967.
- Howard, Neil. "Freedom and Development in Historical Perspective: A Comparison of Gandhi and Fanon's Approaches to Liberation." *Journal of Pan-African Studies*, vol. 4, no. 7, 2011, pp. 94–108.
- Howard, Ryan. *Critique of Nonviolent Politics: From Mahatma Gandhi to the Anti-Nuclear Movement*. Published online, 2002.
- Kebede, Messay. "The Rehabilitation of Violence and the Violence of Rehabilitation." *Journal of Black Studies*, vol. 31, no. 5, 2001, pp. 539–562. <https://doi.org/10.1177/002193470103100502>
- Landry, Jean-Michel. "La Violence Symbolique Chez Bourdieu." *Aspects Sociologiques*, vol. 13, no. 1, 2006, pp. 85–92.
- Mantena, Karuna. "Another Realism: The Politics of Gandhian Nonviolence." *American Political Science Review*, vol. 106, no. 2, 2012, pp. 455–470. <https://doi.org/10.1017/s000305541200010x>
- Martin, Brian. "A Review Article of *Critique of Nonviolent Politics: From Mahatma Gandhi to the Anti-Nuclear Movement* by Howard Ryan." *Pacifica Review*, vol. 9, no. 1, 1997, pp. 83–91.
- Parris, LaRose T. "Frantz Fanon: Existentialist, Dialectician, and Revolutionary." *The Journal of Pan-African Studies*, vol. 4, no. 7, 2011, pp. 4–23.
- Seshadri-Crooks, Kalpana. "I Am a Master: Terrorism, Masculinity, and Political Violence in Frantz Fanon." *Parallax*, vol. 8, no. 2, Apr. 2002, pp. 84–98. <https://doi.org/10.1080/13534640210130430>

Sonnleitner, Michael W. "Of Logic and Liberation." *Journal of Black Studies*, vol. 17, no. 3, 1987, pp. 287–304. <https://doi.org/10.1177/002193478701700302>

Winter, Yves. *Machiavelli and the Orders of Violence*. Cambridge University Press, 2018.

Zeilig, Leo. "Frantz Fanon, Une Vie Révolutionnaire." *CONTRETEMPS*, 12 Dec. 2016, <https://www.contretemps.eu/zeilig-fanon-vie-revolutionnaire/>



Amor Mundi: Political Metamorphosis of Love

JULIAN MINICONI

This paper investigates the place of love within Hannah Arendt's work. Love has been an understudied topic in Arendtian literature, with existing contributions choosing to argue for limited domains of love in Arendt's works, such as the ethico-religious sphere. Distinctively, this paper claims that Arendtian love can be understood politically; more specifically it argues that love can be understood as a political principle inasmuch as it undergoes a process of metamorphosis, actualizing love in the political realm by overcoming a series of paradoxes, each corresponding to one of the three forms of love excavated here: love as private sentiment contained in the ethico-religious sphere, love as a public drive for *praxis* and love as a political principle of the revolutionary spirit spatially actualized as the life of the *polis*. The method of metamorphosis is an interpretive philosophical method aimed at revealing the essence of a given phenomenon through the interplay of the forms it bears. This is similar to the Nietzschean method of 'philosophizing with a hammer' in that it mobilizes discursive agonistic measures, here confronting love with its paradoxes so as to discover a new role/value for love. Ultimately, this paper wishes to propose a new understanding of love, widening the scope of its contribution as a participation in thinking political theory 'otherwise,' concluding with potential applications of love as a political principle in re-imagining the concept of authority and the governance of emerging technology.

Keywords: Hannah Arendt, political theory, political love, metamorphosis, revolutionary spirit

“Now I bid you lose me and find yourselves; and only when you have all denied
me will I return to you.

Verily, my brothers, with different eyes shall I then seek my lost ones; with a
different love shall I then love you.”

– Nietzsche, *Thus Spoke Zarathustra*: First Part, 190

This paper investigates the place of love within Hannah Arendt’s work. Love arguably exerts an important conceptual and emotional influence on Arendt’s overall political theory. It is the central topic of her first published work, *Love and Saint Augustine*, and remains, as we will try to demonstrate, a major concern throughout her other works. Love is a complex affair, if not a fundamentally paradoxical one. In keeping with Arendt’s approach, love demands courage from the political theorist in the attempt to defend or rebuke its ideal: the choice is of human consequence, of potentially fatal deceit or abasement. This paper will seek to salvage love from this tragic dilemma, following Arendt’s lead by diving headfirst into its paradoxes, which will be reflected in the structure of our argument. It will also mobilize Arendt’s works across the board, paying special attention to her book on St. Augustine. Each section of this paper will benefit from the insights of those who appear to be among her most revered authors: Machiavelli, Rousseau, Nietzsche, and Kierkegaard, each contributing in their own respect to our inquiry.

The topic of Arendtian love has been surprisingly understudied. Existing literature on Arendt’s concept of love is mostly focused on interpreting her dissertation on St. Augustine, seeking to recover a “public” understanding of Arendtian love through the notion of friendship (Chiba, 505-35). Distinctively, this paper chooses to approach Arendtian love throughout the whole architecture of her work in order to reconstruct what seems to be her mostly concealed phenomenological concern for love. The aim of this paper is to excavate the notion of love from Arendt’s corpus and try to unfold its political significance by demonstrating its affinity with other major trends of Arendt’s works, such as the public/private dichotomy, freedom, violence and the revolutionary spirit. As such, this paper contributes to both the limited literature on Arendtian love but also to the overall literature on Arendt by providing a novel outlook on her political theory. Ultimately, this paper wishes to propose a new understanding of love, widening the scope of its contribution as a participation in thinking political theory ‘otherwise’ and concluding

with potential applications of the political principle in reimagining the concept of authority and the governance of emerging technology.

This paper will unfold Arendtian love through the image of metamorphosis, claiming that love can be understood as a political principle inasmuch as it undergoes a process of transformation: initially contained within *appetitus* as a private sentiment, it is transformed into a drive for *praxis* by claiming worldliness. Spreading in the public as a drive for *praxis*, love takes a chaotic form, resting on an emotional mode of association and engendering violence. The political actualization of love is made possible by its second metamorphosis, operationalized by the crystallization of love within the revolutionary spirit, claiming a world-building capacity by virtue of the spatial actualization of its drive for novelty. Ultimately, this metamorphosis of love would present love as a political virtue of *la vita activa* and a political principle of the revolutionary spirit.

The method of metamorphosis is an interpretive philosophical method aimed at revealing the essence of a given phenomenon through the interplay of the forms it bears. This is similar to the Nietzschean method of ‘philosophizing with a hammer’ in that it mobilizes discursive agonistic measures, here confronting love with its paradoxes so as to discover a new role/value for love. This method is here both a theoretical exercise and empirical narrative aiming at representing the journey of love within the *polis*. That is, each section is committed to confronting a given theoretical form of love to its empirical appearance. As such, two levels of metamorphosis could be identified; a first level laying out the claims one can make about love through theoretical exercise, a second evaluating the existence of love as it acts in the world it claims. These two levels interact with each other dynamically. The intuitions of the former level are checked by the latter, which in turn informs new intuitions that are yet again confronted with their empirical existence, until one reaches a suitable conception of love, one that can hold its existence in the world, specifically here the *polis*, without compromising itself.

1st Metamorphosis of Love: Claiming the World

“Great Heaven! The prince says that beauty saves the world!”

– Fyodor Dostoyevsky, *The Idiot*

This first section will articulate the first metamorphosis of love, from a

private sentiment to a public drive for *praxis*, arguing that such a transformation operates when love comes to claim worldliness as a consequence of its outburst in the public, ultimately playing an important role in Arendtian freedom.

In her dissertation, *Love and Saint Augustine*, Arendt exposes the Augustinian theory of love. Her task in what would be her first published book was analytical; that is, Arendt was committed to laying down the contradictions in Augustine's theory, hoping to discover within them what would be a hidden truth. She sought, specifically, to reconstruct the Christian creed of loving one's neighbour like oneself through Augustine's definition of love as *appetitus* (7). The concept of *appetitus* designates a "craving" towards a given object, with the object itself determining the value of such love as craving (9). "Wrong love," or *cupiditas*, consists of an attachment to the mundane world informed by an aversion to death, whereas "right love," or *caritas*, is concerned with "eternity and the absolute future" (17). This first distinction within the concept of love leads us to our first paradox in establishing love as a political virtue, namely that virtue seems to be exclusively located in the unworldly *caritas*, whereas politics and worldliness would be located in deficient *cupiditas*. Arendt helps us solve that first paradox by ascribing to love a 'derivative' order: love derives from a concern for eternity, itself sourced in the faculty of remembrance, and then unfolds in the world. We come to understand 'loving our neighbour' as a function of that derivative: "He occupies this place because he is like myself; he, too, can 'enjoy God' (*Deo frui*). He is my neighbour only insofar as he has entered into the same relationship with God or the absolute future" (42, 48).

This first inquiry dwells in the nature of love itself and lays out the tenet of our first paradox, the place of love in the public/private dichotomy and ultimately its relationship to freedom, which is also to be found in *The Human Condition* wherein Arendt states that: "Because of its inherent worldlessness, love can only become false and perverted when it is used for political purposes such as the change or salvation of the world" (51-52). That is, love approached as *appetitus*, because it is sourced in a relationship towards "eternity", something fundamentally unintelligible and unattainable, cannot be reflected under the light of the public realm. As such, its claim to freedom within Arendt's political theory is problematic (Chiba, 505-35)¹, inasmuch as freedom is understood as a public good:

¹ Shin Chiba, "Hannah Arendt on Love and the Political: Love, Friendship and Citizenship," The previous derivative of friendship having already been explored in the literature, this paper will not comment

“Freedom needed ... the company of other men who were in the same state, and it needed a common public space to meet them ... into which each of the free men could insert himself by word and deed” (Arendt, *The Human Condition*, 147).

This first predicament concerning the fitness of love as *appetitus* with freedom and worldliness leads us to formulate the first metamorphosis of the concept of love, enabling love to make its appearance in the public realm as a driving force of free *praxis*, consistent with Arendt’s derivative approach to love. It is through that first metamorphosis that love comes to claim worldliness while retaining the motive of ‘eternity.’ Love hence takes a new form, that of a drive for action. This transformation follows from an interpretation of Arendt’s Jesus of Nazareth and Heinrich Blücher’s non-spiritual account of Jesus’s deeds and sayings. In *Between Past and Future*, Arendt chooses to close her chapter on freedom with a surprising account of miracles. She relates the phenomena of miracles to the concept of birth and freedom to demonstrate the world-building capacity of free action in interrupting “some natural series of events, of some automatic processes, in whose context they constitute the wholly unexpected” (Arendt, *Between Past and Future*, 166). Miracles, viewed as a human faculty for action, function in the political realm as unpredictable interrupting events, making the creativity of the deed or word proportional to its world building capacity (169). This approach to miracles is similar to Blücher’s interpretation of Jesus of Nazareth, who he understood “as someone who had something absolutely new and amazing to say, who brought ideas into the world without which we would not be able to live, eternal ideas to be eternally considered” (2). The inspiration of eternity makes the praxistical love of Jesus properly apolitical, even though it carries an ethical message that manages to resonate within the public realm and that functions as a drive for free action: “This hope that nobody is left out, that nobody will ever be left out, is the hope that he brought into the world, and it is the hope by which all free men still live whether they know it or not,” writes Blücher (6). However, it is a crucial step for imagining the world-building capacity of love and its capacity to carry an inspiration for free action into the public realm, hence avoiding the first paradox of the yet-to-be political principle of love, namely that it could not be carried out in the public realm because of its “inherent wordlessness” (Arendt, *The Human Condition*, 51-2), but we have yet to treat how this love unfolds in the

further on it, except that it could be interesting to approach political friendship or partisanship, in the noble sense, as a form of political love in its own respect..

public realm.

The Errance of Love as a Drive for Praxis: Violence and Totalization

“A crown was given unto him:
and he went forth conquering,
and to conquer”

– *The Book of Revelation*, Rev. 6:2²

Now that love has undergone its first metamorphosis, initially apprehended as an *appetitus* before being transformed into a drive for *praxis* by claiming worldliness, it is yet again afflicted by two paradoxes obscuring its way toward political actualization, which this paper will presently tackle. Firstly, as a driving inspiration of *la vita activa*, this concept of love seems to be particularly vulnerable to Arendtian violence, an argument which will be supported by Machiavelli’s own insights on the unstable nature of love. Secondly, and relatedly, love as it is understood here still seems to rely heavily on an emotional mode of association that would risk falling prey to dogmatism and elitism as Arendt saw in her critique of the 1960s student protest movement in *On Violence*. This second argument will be supported by Rousseau’s theory of the general will, which we will identify as the theoretical wedge leading us to formulate our second metamorphosis and definition of love. This will lead to love ultimately acquiring its political character.

Love as a drive for free action appears in the public realm chaotically, consistently with the derivative relationship between its world-building capacity and eternity/infinity, constituting the inspiration of its *praxis*. This problem was outlined by Machiavelli, who contended that a virtue of the Republic was its capacity to regulate such effusions through the conflictual balancing of good institutions, which would channel them into desirable political outcomes for “men never good except out of necessity, but where choices are abundant and unlimited freedom is the norm, everything immediately becomes confused and disorderly” (Machiavelli, *Discourses*, 28). This republican virtue could ‘regulate’ love if we consider the drive inspired by love to be similar to the drive inspired by any other

2. No argumentative parallel between the figure of Jesus of Nazareth and the cavalier of the Apocalypse is meant here, although one could establish a relationship between them as two faces of the Janus-like articulation of freedom containing both a loving and conquering ego.

motive. However, for Machiavelli, love among humans or as a principle for action cannot be relied on politically as “men in most instances deceive themselves in the matter of the love they judge that someone bears for them; nor can you ever be sure of this love unless you test it” (261). The difficulty of equating love with a stable drive for *praxis* was also reflected on by Arendt, who in mentioning the spatial dimension of freedom claimed: “To be sure it may still dwell in men’s hearts as desire or will or hope or yearnings, but the human heart, as we all know, is a very dark place, and whatever goes on in its obscurity can hardly be called a demonstrable fact” (Arendt, *Between Past and Future*, 147). This quote could be interpreted along the lines of the first paradox of political love, its unworldliness, but can also be understood in terms of reliability, similarly to Machiavelli’s insights. That is, it is difficult, and even dangerous, to assign a loving intention to a political action *a priori*. This predicament leads love as a drive for *praxis* to unfold inadequately in the political realm, either embracing a chaotic expression and leaving the door open for violence or attaching itself to a certain object and elevating it as a dogma, with both preventing its ‘adequate’ political actualization.

We will now examine this predicament closer, informed by Arendt’s treatment of the 1960s student protests in *On Violence*. Specifically, we will argue that love as a drive for *praxis* risks falling prey to various captures, namely from dogma, violence and elites, constituting its second and related paradox. According to Arendt, the student movement of the 1960s was plagued by its commitment to the relic of “progress” (Arendt, *On Violence*, 30). Arendt considered that the movement’s attachment to the idea of ‘progress’ led to both the capture of the movement’s claims by elites but also the legitimization of violence: “If we look on history in terms of a continuous chronological process, whose progress, moreover, is inevitable, violence in the shape of war and revolution may appear to constitute the only possible interruption” (30). If, as Arendt claims, “To be sure, every revolutionary movement has been led by the disinterested, who were motivated by compassion or by a passion for justice,” often this emotional force, also contained in love as drive for *praxis*, has been captured and instrumentalized to serve the interest or purpose of a particular class and dogma (24).

Such a mode of association can only arise by following the direction of the ‘general will,’ which contains both the will of all, and an overarching collective spirit geared towards the realization of freedom: “From public lights result the union of the intellect and the will in the “social body,” from there derives

the exact concord of parties, and, ultimately, the greatest strength of the whole” (“Alors des lumières publiques résulte l’union de l’entendement et de la volonté dans le corps social; de là l’exact concours des parties, et enfin la plus grande force du tout.”; Rousseau, 96)³. That is, the ‘general will’ is both the expression of a collective will and, most importantly, the expression of a sentiment driving a public understanding of freedom as *praxis*. This element of the ‘general will’ fits our definition of love as a drive for *praxis* particularly well, which seems to be also reflected in Rousseau’s overall articulation of the body politic wherein the general will, equated with the legislative function consistently with Rousseau’s system, resides in the heart of the body (153).

However, the ‘general will’ is a notoriously vulnerable concept, suffering from an authoritarian inclination already present in Rousseau’s *Social Contract* and being highly manipulable. Rousseau famously argued in *Du Contrat Social* that any person that deliberately corrupted, sabotaged or harmed the general will, and was judged guilty of that, would become the enemy of the general will and liable to death: “he must be cut off by exile as a violator of the compact, or by death as a public enemy; for such an enemy is not a moral person, he is a man” (“il en doit être retranché par l’exil comme infracteur du pacte, ou par la mort comme ennemi public; car un tel ennemi n’est pas une personne morale, c’est un homme [...]”; 91). The revolutionary life of Robespierre and his relationship to the Parisian crowd provide a prime example of this tendency of the general will. Robespierre, who once galvanized the Parisian crowd as the embodiment of the general will and was admired as the “Incorruptible” during the most violent days of the revolution, found himself only a few weeks later on the scaffold in a pathetic state, mocked, in a shocking contrast, by the mob cursing him as their “Tyrant” (Jones, 692). The effusion of love contained in the ‘general will’ and energizing the Parisian crowd, the same passions that Arendt identifies as the cause of the French Revolution’s failure, turned against the people that tried to capture it (Arendt, *On Revolution*, 214). It also failed to provide love with its political actualization, taking a form described by Arendt in *The Human Condition* as “perhaps the most powerful of all antipolitical human forces” (242).

Therefore, love as a drive for *praxis* needs to undergo a second metamorphosis, inspiring itself from Rousseau’s general will, in that it can appear as a motive for a collective exercise of freedom, acting both as a drive for *praxis* and as a

³ Translations for this text were done by me.

social glue. However, the social and the political, if we wish to be consistent with Arendt, cannot be equated. Moreover, while the social is certainly blown away by time, explaining the vulnerability of passions during the French Revolution, the political, because it occupies a space for itself, deflects or interrupts time. This is how love needs to be actualized, as a spatial point of reference for freedom that crystallizes its drive for *praxis*; only then can it claim a political existence. Initially arising from the dark realm of the human heart as a craving, then unleashed in the public as a drive for *praxis*, love now needs to be crystallized so as to prevent its violent wandering and allow for its political actualization. This paper will argue that such a crystallization operates at the birth of a new political project, ultimately taking the form of Arendt's actualized revolutionary spirit. The crystal of political love, equated with the actualized revolutionary spirit, constitutes a world in-itself and for-itself, an argument which will be supported by Arendt's concept of space and time and ultimately Kierkegaard's concept of the 'leap'.

2nd Metamorphosis: Crystallization of Love in the Revolutionary Spirit

“I love my native city more than my own soul”

– Niccolò Machiavelli, *The Letters of Machiavelli*

Arendt defines the central idea of a revolution as “the foundation of freedom, that is, the foundation of a body politic which guarantees the space where freedom can appear” (*On Revolution*, 116), and which can act as a point of remembrance of the human capacity for novelty (215). Both of these elements articulate the revolutionary spirit, which is carried out initially as a drive for *praxis* geared towards the actualization of freedom in the *polis*, a *praxis* containing the same modal character as love as a drive for *praxis*, but ultimately differentiated by its political end which aims at its spatial crystallization. This last element guarantees the existence of love as a world-building drive by inserting its reality within space and time. Arendt, building on Franz Kafka's parable of time, considered that “Seen from the viewpoint of man, who always lives in the interval between past and future, time is not a continuum ... it is broken in the middle” (Arendt, *Between Past and Future*, 10). That is, the forces of time converging at the spatial position of the individual are interrupted by space itself, claiming its own reality at the intersection of past and future. Similarly, the actualized revolutionary spirit denotes an interruption in history, acting as a point of reference itself for remem-

brance and inspiring novelty: an action crystallized in history and spatially actualized. Love infused in the revolutionary spirit creates a breach in the continuum of time, which attaches to the ‘founding’ act of freedom a world-building faculty. It is precisely a point of reference for remembrance for the purpose of novelty; both *appetitus* and the chaotic emotional love as a drive for *praxis* are contained in the crystal, which reflects time and answers fate with will. A revolt against time best expressed by the poets, for whom the drive of love can never be forgotten:

“The winds must come from somewhere when they blow,
There must be reasons why the leaves decay;
Time will say nothing but I told you so.
Perhaps the roses really want to grow,
The vision seriously intends to stay;
If I could tell you I would let you know.” (Auden, 10-15)

Auden’s verses reflect the paradoxical struggle of freedom with time, of seeking to derive meaning from what determines us and receiving back mere silence. Alone, faced with its world-building capacity, freedom becomes “dizzy” and entangled in an incommensurable realm of possibilities (Kierkegaard, *The Concept of Anxiety*, 61). What love inspires is the “leap” (Kierkegaard and Bretall, 200), which is both its starting point, which prevents this paper’s argument from pretending to be systematic, and the object of remembrance of its crystallized form, or spatial and political actualization.⁴ It supports the world-building drive of freedom by being a living example of it: the ultimate form of love constitutes the life of the *polis*. Therefore, the spatial form of political love requires a suitable ‘receptacle.’ The world-building capacity given to love, and most importantly its active role in that process, is not sufficiently covered by a document such as a constitution, which merely reflects a founding deliberation process but must consistently be embodied in active sovereign assemblies. So that the revolutionary spirit may never die, its love must be crystallized in its children, ultimately giving love its full political actualization: “[Thomas Jefferson’s] exposition of the ward system always began with a reminder of how the vigour given to our revolution in its commencement was due to the ‘little republics’, how they had ‘thrown the whole nation into energetic action’ ... there was not an individual in their States whose body was not thrown with all its momentum into action” (Arendt, *On Rev-*

4. An interesting theoretical pathway could be made here relating remembrance within political love to the corollary feelings of melancholia and nostalgia, analyzing for their own phenomenology within the political realm.

olution, 242-243). This is also how we can truly understand Machiavelli's claim,⁵ which seemed at odds with his treatment of love in the *Discourses on Livy*; it is when love has overcome itself, crystallized "beyond good and evil",⁶ that it can claim to be the life of the *polis* (Nietzsche, 444).

Conclusion

Love, here actualized politically as the life of the *polis* and an active point for remembrance, crystallizes both its inspiration for novelty and preservation. One of the properties of this political form of love is that it is self-contained; that is, the force of love as a political principle rests in its ability to overcome division, for example, partisan or ideological division, by holding fast to its essential ambiguity. This characteristic of the political principle of love seems to be also reflected linguistically in the founding act of the *polis*, according to Arendt: "Terminologically speaking, the effort to recapture the lost spirit of revolution must, to a certain extent, consist in the attempt at thinking together and combining meaningfully what our present vocabulary presents to us in terms of opposition and contradiction" (Arendt, *On Revolution*, 215-216). This attribute is also a virtue of the principle inasmuch as it does not erase or destroy division but simply gives motion to the life of the *polis*, including its internal dynamics. This force should be wisely used so as to not overflow, or conversely famish the *polis*, two directions from which we could partially derive the frontline of strictly political ideological divisions, such as conservative and progressive trends. The ambiguity of our political vocabulary overall was denoted by political theorists as symptomatic to the complex affairs of government, among them Michael Oakeshott who premised his theory of modern government in the Western world on the fact of the necessary division of our political language and conduct as a consequence of this ambiguity and ambivalence (11-12). Comparatively to our articulation of the political principle of love which acknowledges both the severity of preservation and the thrilling of novelty, such a premise is not mobilized to be overcome but merely to advocate for a 'golden mean' understanding of government such a balance favoring, in Oakeshott's political theory, a skeptical leaning. Future attempts to articulate the political principle of love, inasmuch as they strive to think political theory 'otherwise' by bringing forward a new outlook on love's essential

⁵"I love my native city more than my own soul."

⁶ Full quote: "What is done out of love always occurs beyond good and evil."

potential, should explore the principle's relationship to authority and its potential for acting as a principle guiding governance. Specifically, the political principle of love could be an interesting theoretical pathway for thinking about the governance of emerging technologies, seen as the next playground for revolutionary resolve. Besieged by its own paradoxes in its attempt to navigate authority's own domain, and technology's phenomenology,⁷ perhaps the next metamorphosis of love would need to start as a linguistic effort, consistently with its creative drive, aimed at constructing a word or concept that could capture the meaning of this new political principle.

7. In *The Question Concerning Technology*, Heidegger's reflections on the essence of technology are a fertile ground of investigation in that regard, particularly the idea that in order to reveal the, partial, "world-saving" essence of technology: "Because the essence of technology is nothing technological, essential reflection upon technology and decisive confrontation with it must happen in a realm that is, on the one hand, akin to the essence of technology and, on the other, fundamentally different from it" (14).

Works Cited

- Arendt, Hannah. *Between Past and Future: Eight Exercises in Political Thought*. Penguin, 1977.
- . *Love and Saint Augustine*. Chicago University Press, 1996.
- . *On Revolution*. Penguin Books, 1963.
- . *On Violence*. Harcourt Brace Jovanovich, 1970.
- . *The Human Condition*. The University of Chicago Press, 2018.
- Auden, W. H. “If I Could Tell You.” Allpoetry.com, 1940.
- Blücher, Heinrich. “Jesus of Nazareth: Two Lectures by Heinrich Blücher New School for Social Research.” Blücher Archive, 1954.
- Chiba, Shin. “Hannah Arendt on Love and the Political: Love, Friendship, and Citizen ship.” *The Review of Politics*, vol. 57, no. 3, [University of Notre Dame du Lac on behalf of Review of Politics, Cambridge University Press], 1995, pp. 505–35, <https://doi.org/10.2307/1408599>. JSTOR.
- Dostoyevsky, Fyodor. “The Idiot.” 1869. *Books Abroad*, vol. 17, no. 3, Books Abroad, 1943, <https://doi.org/10.2307/40083533>.
- Heidegger, Martin. *The Question Concerning Technology*. 1954. Martin Heidgger: Basic Writings, David Farrell Krell, 1977, www2.hawaii.edu/~freeman/courses/phil394/The%20Question%20Concerning%20Technology.pdf.
- Jones, Colin. “The Overthrow of Maximilien Robespierre and the ‘Indifference’ of the People.” *The American Historical Review*, vol. 119, no. 3, June 2014, pp. 689–713, <https://doi.org/10.1093/ahr/119.3.689>.
- Kierkegaard, Søren, and Robert W. Bretall. *A Kierkegaard Anthology*. Princeton University Press, 1946.
- Kierkegaard, Søren. *The Concept of Anxiety: A Simple Psychologically Orienting Deliberation on the Dogmatic Issue of Hereditary Sin*. Princeton University Press, 1980.
- Machiavelli, Niccolo. *Discourses on Livy*. Oxford University Press, 2003.
- . *The Letters of Machiavelli*. Translated by Gilbert, Capricorn Books, 1961.
- Nietzsche, Friedrich Wilhelm. *The Portable Nietzsche*. Edited & translated by Walter Kaufmann, Penguin Books, 1982.
- Oakeshott, Michael, and Timothy Fuller. *The Politics of Faith and the Politics of Scepticism*. Yale University Press, 1996.
- Rousseau, Jean-Jacques. *Du Contrat Social*. Editions Sociales, Les Classiques du Peuple, 1971.
- “The Book of Revelation (King James Version).” *Early Christian Writings*, p. Rev. 6: 2.



Fanon in the Consulting Room: Colonialism, Racism, and the Urgent Need for a More Inclu- sive Mental Healthcare System

ANAHAT GHUMAN

Conversations surrounding mental health are becoming more inclusive; however, the lived experiences of marginalized communities continue to be underrepresented, underserved, and mistreated by the current mental healthcare system. This essay aims to analyze the current mental health crisis that disproportionately impacts people of colour (POC) in the West and postcolonial communities through the lens of Frantz Fanon's nuanced ethno-psychiatric framework. I argue that the 'psychiatric social therapy' advanced by Fanon, along with its psycho-philosophical humanistic underpinnings, helps explain the ongoing mental health crisis and provides impetus for significant improvements to the current mental healthcare system. Ultimately, this essay criticizes the Western approach to mental healthcare as endorsed by the American Psychiatric Association (APA), as it perpetuates injustice to POC in the Western context as well as diverse postcolonial communities across the globe where this approach has been adopted. I will be endorsing Fanon's proposed outlook with reference to two distinct groups: POC in the West (North America) and peoples from postcolonial countries, with a specific example of South Africa. Fanon asserts his theory largely in reference to the experience of being a colonial subject, focusing on the intersection of blackness and colonialism in the Afro-Caribbean context. However, in this essay, I will expand this frame of reference to BIPOC communities in the West and analyze the African and South Asian postcolonial experience vis-à-vis mental healthcare.

Keywords: mental health, healthcare, Frantz Fanon, colonialism, racism, culture, psychology, psychiatry, psychotherapy, neo-colonialism, American Psychiatric Association, Diagnostic and Statistical Manual (DSM).

Introduction

While conversations around mental health are becoming more inclusive, the lived experiences of marginalized communities continue to be underrepresented, underserved and mistreated within the current mental healthcare system. This essay aims to analyze the current mental health crisis that disproportionately impacts people of colour (POC) in the West and postcolonial communities through the lens of Frantz Fanon's nuanced ethno-psychiatric framework. I argue that the 'psychiatric social therapy' advanced by Fanon and its psycho-philosophical humanistic underpinnings helps explain the ongoing mental health crisis and provides impetus for significant improvements to the current mental healthcare system. Ultimately, this essay criticizes the Western approach to mental healthcare as endorsed by the American Psychiatric Association (APA), as it perpetuates injustice to POC in the Western context as well as diverse postcolonial communities across the globe where this approach has been adopted.

While Fanon's theory centres on the colonial subject, primarily within the Afro-Caribbean context, I will be endorsing his proposed outlook by significantly expanding its application. Specifically, I will examine the mental healthcare experiences of POC in the West (North America) and peoples in postcolonial countries—with a narrow focus on the South African case—demonstrating the broader relevance of Fanon's insights.

This essay is divided into three main sections for clarity. The first section will outline the nature of the current mental health crisis worldwide and its disproportionate impacts on the groups of interest, using the example of the DSM's Post Traumatic Stress Disorder. Next, the second section will focus on Fanon's framework for psychiatric conditions and care, underpinned by his philosophical and psychiatric work. Lastly, the third section will argue that gaps in the current approach to mental health can be filled by realistically considering and applying Fanon's suggestions regarding psychiatric social therapy.

Ongoing Mental Health Crisis: Failing of the Western Mental Health Framework

By World Health Organization estimates, 1 in 8 individuals across the globe live with a mental disorder ("Mental Health Disorders"), with one meta-analytical study suggesting that about 14.3% of deaths worldwide may be attributed to mental disorders (Walker et al.). Because globalization and the digital age have

increased access to information as well as connectivity, there has been a significant rise in mental health activism, leading to enhanced societal sensitivity toward psychological disorders (Tribe and Melliush; Okasha; Herrera-Peco et al.). However, there remains a steady upward trend in mental disorder rates, with the 2020 COVID pandemic exacerbating rates of mental illness, suicide, and substance use among various demographics (Panchal et al.). Thus, there is clear evidence of an ongoing global mental health crisis, with detailed and widely available statistics for North America, making it crucial to understand the impact on minorities to ensure they are able to access appropriate means of treatment and care.

An essential component of my criticism of the mental healthcare framework is the American Psychological Association's Diagnostic and Statistical Manual (DSM), given that it is touted to be the 'bible' of mental healthcare and yields a key influence on medical care. The DSM-5 reflects a paradigm shift in mental healthcare as it bolsters an atheoretical classificatory approach to mental health and is characterized by its "technical language, diagnostic categories, elaborate classification systems, and empirical support" (Jacob et al.). This, in itself, warrants much criticism when we consider that alternative and historically marginalized theoretical representations, distinct from the dominant biomedical approach, have consistently acknowledged social, cultural, racial, and ethnic dimensions of community mental health (Jacob et al.), insights that are often highlighted by fields such as medical anthropology. Its underlying biomedical framework means that the locus of deviant behaviours is often conceptualized to be within an individual and their body (specifically, their neurochemistry), resulting in a system that medicalizes personal distress, operates with the assumption of biological causation, and argues for pharmacological treatment as the first line of defense (Jacob et al.). While presented as an atheoretical classification system, the DSM-5's practical application reflects a principally biomedical framework that operates with the assumption of biological causation (Jacob et al.).

It is beneficial to consider one category of the DSM-5 that is pertinent to both our groups of interest: Post-Traumatic Stress Disorder (PTSD). PTSD regards the long-term impact of traumatic events (such as war, sexual assault, natural disasters and so on) on individuals. There are wide-ranging concerns with the current PTSD framework: it pathologizes rational and sometimes important reactions to trauma such as anxiety, hyper-vigilance, and numbing, and the 'post'

in PTSD reflects the inconsideration of ongoing trauma and abuse that many individuals may be facing on a daily basis (Jacob et al.). For POC and women, occurrences are the norm (Vernor; Franklin et al.), yet they receive no specific consideration within the DSM. The DSM shows an ethnocentric bias through its implicit assumption of universal trauma responses and its failure to adopt a trauma-specific lens in healthcare (Kirmayer, *Culture, Context and Experience*). In line with this, we could postulate that an individual routinely facing racism and/or microaggressions may receive the same diagnosis as someone who, for instance, lived through a hurricane years prior. This temporal and categorical disregard negatively impacts POC in the West and ignores systemic colonial intergenerational trauma suffered by those in postcolonial contexts (Mills). The DSM-5 has been heavily critiqued for medicalizing normal human reactions to life events (News; Dyck); this ties into the pathologization of normal reactions of people of color to aversive events, which are often a function of their identity in society.

This is not simply restricted to trauma, as the psychological distress experienced by POC in systematically biased societies, manifested as a myriad of ‘symptoms,’ is often a direct response to systemic oppression (Comas-Díaz et al.; Sue; Williams & Mohammed), which may cease under more equitable conditions, thereby contradicting the DSM-5’s individualistic characterization of disorders, which often fails to account for sociopolitical and cultural contexts (Kirmayer, *Beyond the ‘New Cross-Cultural Psychiatry’*; Marsella; Sue, 15).

Within the North American context, the disproportionate impact on communities of colour is apparent, from standards of diagnosis to quality and investment in care. Minority patients, as compared to White individuals, are less likely to avail the best treatments for depression and anxiety (which are the most common diagnoses affecting people worldwide); accordingly, we may infer this to be a contributing factor to African-Americans terminating treatment prematurely (Sue). Healthcare providers work with ‘priors,’ i.e. preconceived estimates of the likelihood that a patient has a disorder, which are then reevaluated based on interacting with the patient (McGuire & Miranda). In keeping with this, physicians are less willing to diagnose variations in the severity of depression within minorities, which signals the inability to correctly understand the prognosis of minority mental health (McGuire & Miranda). This can be interpreted as both a cause and a consequence of the underrepresentation of minorities in mental disorder prevalence rates (McGuire & Miranda). Specifically, the bias in diagnosis leads to un-

derreporting, which in turn reinforces the bias (McGuire & Miranda). If clinicians consistently fail to recognize and diagnose depression in minority patients, those cases will not be reflected in official statistics (McGuire & Miranda). Additionally, the lack of minorities represented in the data causes less research to be done on the specific mental health issues of those minorities, which in turn causes less training for clinicians, thus making it such that the clinicians have less accurate ‘priors’ regarding those minorities (McGuire & Miranda). Studies have shown a clear bias in the diagnosis of mental health conditions in African-American individuals. This bias manifests in specific diagnostic patterns. For example, they are often overdiagnosed with disorders like schizophrenia, while being underdiagnosed with conditions like depression (Strakowski et al.; Neighbours et al.).

The move to standardize the application and measurement of psychological disorders has resulted in the Western mental health framework being implemented to various degrees in postcolonial countries around the world, including Algeria, South Africa and India (Mills). This implementation is influenced by the APA and the DSM-5 (Mills; Jacob et al.). The limitations of an atheoretical approach such as the DSM’s are acutely apparent in its lack of cultural competency and ethnocentricity and thus, one of the main developments to be incorporated in the DSM-5 are cultural considerations. However, many analyses have found the attempt ineffective, criticizing the DSM for conflating culture with race, ethnicity, or nation and for stereotyping groups in terms of fixed cultural traits (Bredström). Susan Kriegler, a psychologist working and teaching in South Africa, has noted how the application of DSM criteria in child psychiatry has immense potential to stigmatize young children who are labelled according to the foregoing criteria (Kriegler, 604). Additionally, she highlights two primary concerns. First, she explores how South Africa reflects the transgenerational impact of trauma, which does not lend itself to pathologization with reference to any existing DSM criteria (606). Another important consideration made in her paper is the inequitable access to mental health resources in South Africa, which may potentially increase inequity with selective (over)treatment of affluent families that can afford it and undertreatment and underdiagnosis of those who are not as socioeconomically privileged (607).

There also exists evidence that biogenetic explanations lead to essentialist thinking, wherein those labelled as having a disorder are seen to share an immutable, defining nature that makes them “essentially” different (Kvaale et al.).

Historically, the phenomenon of essentialism has been associated with racism as it is a mechanism for enforcing and reinforcing stereotypes about groups (Mandalaywala et al.). Moreover, the aforementioned biogenetic explanations have been associated with labelling those with disorders as dangerous (Mandalaywala et al.). An alarming intersection can be found when we consider the enduring and harmful racist stereotype of aggression and danger attributed to Black men, evidencing another potentially problematic aspect of the current model (Mandalaywala et al.).

Fanon's Existential Psychiatric Framework

Frantz Fanon was a Martinican psychiatrist and activist working under the French colonial order, stationed in Algeria and Tunisia. His seminal works—*Black Skin, White Masks* and *The Wretched of the Earth*—isolate and describe various concepts that highlight the psychological impacts of colonialism and racism. In order to understand and appreciate the value of his suggestions in the field of psychiatry and psychotherapy, it is essential to consider the aforementioned psychological impacts of colonialism and racism, as well as the phenomenological and existential frameworks that underlie and help explain these impacts in Fanon's work. As was mentioned above, I believe that, although Fanon focuses on the Afro-Caribbean context, many of his concepts can be extended and applied to other racial groups, as racism and colonialism have manifested themselves across diverse racial contexts. Many imperial powers have utilized and enforced similar ideologies and policies to legitimize their oppression of Indigenous peoples in the lands they wish to conquer, often resorting to essentialist thinking that reduces entire groups to fixed, unchanging traits to justify their subjugation. The shared rhetoric of 'savages' or 'uncivilized people' exemplifies this essentialist approach, used as a legitimizing force by the United States of America and Canada against indigenous populations and by British Imperialism in Asia and Africa (Mackay et al.; Rahman).

At the centre of *Black Skin, White Masks* is Fanon's goal: to analyze and deconstruct the "massive psychoexistential complex" created by colonization and racialization (5). Existential psychology is a prominent branch of contemporary psychology that informs a form of therapy known as existential therapy. It comprises 5 tenets, as summarized by Vos et al.:

1. Human beings orient themselves to, and have a need for, meaning and

purpose.

2. Human beings have a capacity for freedom and choice, and function most effectively when they actualize this potential and take responsibility for their lives.
3. Human beings will inevitably face limitations and challenges in their lives, and function most effectively when they face—rather than avoid or deny—these givens.
4. The subjective, phenomenological flow of the individual's experiencing—including all senses, both negative and positive experiences—is a key aspect of being human, and therefore a central focus for psychotherapeutic work
5. Human experiencing is fundamentally interrelated with—rather than separate from—the experiencing of other human beings and with its world.

With these tenets in mind, it is not difficult to understand why Fanon would appeal and engage with such a theoretical framework. He was working in a context wherein he faced not only his but also his oppressed patients' struggle to make meaning and survive under a regime that curtailed their freedom and discriminated against them based on their race and ethnicity. He can also be seen in discourse with Jean-Paul Sartre, one of the primary proponents of existentialism, as he considers and replies to many of Sartre's positions in *Black Skin, White Masks*. In *Black Skin, White Masks*, Fanon explores objectification and epidermalization, highlighting how these psychological processes profoundly affect POC, creating a psychic experience distinct from white individuals'. According to Fanon, Black people face objectification due to the economic greed of white groups that have historically benefited from their racialization and subjugation (a conception we may extend to all POC, especially in cases of colonization) (4). This objectification is further entrenched and internalized through 'epidermalization,' a process that profoundly impacts the psyche of racialized individuals (Fanon, 84). These mutually reinforcing processes, rooted in racial and colonial trauma, highlight the need for a framework that acknowledges the structural impact of these experiences on an individual's worldview.

To explore what Fanon means by objectification, I would like to consider the concept of the 'Look' as described by Sartre, which is applicable to any situation where we engage with the Other (i.e. anyone who is not us) in the world.

According to Sartre, human reality consists in being a nihilating consciousness that works to categorize and rationalize the world around us. We view the world around us relative to our projects or goals, whereby we label things/objects as useful or not (Sartre, 193-194). This changes when we encounter the Other (Sartre, 270). We know the Other observes the world around him in a manner similar to ours, hence interrupting our attempt to reduce them to a mere object—what Sartre calls “nihilation” —the reduction of the Other to an object, devoid of their own subjectivity and freedom. This is because their subjectivity, like our own, is inaccessible to us: we cannot fully grasp how they perceive or interpret the world (Sartre, 270). Yet, at the same time, when we become aware that we ourselves are being perceived by the Other, we experience a reversal: instead of being the one who looks, we become the one who is looked at. In that moment, we do not access the Other’s subjectivity, but rather feel the fact of being seen by it—leading to our own experience of being objectified in their gaze (Sartre, 193–194). This objectification means that we view ourselves as we would view a rock because we are robbed of our transcendence, or our ‘being-for-itself’—our conscious awareness, freedom, and ability to project ourselves into the future—and rather become a ‘being-in-itself’—a fixed, determined object, devoid of consciousness and agency (Sartre, 193-194). However, Sartre also outlines how the same process of objectification works in reverse: we, too, incite the feelings of objectification and shame in the Other, that is, the recognition of the fact that “I am indeed that object which the Other is looking at and judging” (Sartre, 199).

While Sartre’s analysis of the Look provides a foundation for understanding objectification, Fanon reveals the crucial ways in which this process is uniquely intensified and distorted for Black people. Fanon highlights how existing in the world as a Black person fundamentally robs him of this double-sided process where only the Black man is objectified and racialized. The existence of a racialized individual is filtered through the lens of toxic historical racial stereotypes, impacting their movement in an attempt to determine it extrinsically (Fanon, 5). This means that their experiences and opportunities are shaped by pre-existing biases and prejudices, limiting their freedom to move and define themselves on their own terms (Fanon, 82). They are constantly being judged and categorized based on external perceptions rather than their own individuality (Fanon, 87).

The racialized individual hence loses a characteristic that Sartre takes as an assumption—the characteristic of being a body or of being-in-the-world.

Fanon's concept of the racial epidermal schema highlights how the internalization of external expectations restricts the individual's movement, both physically and socially, contrasting with Sartre's vision of a freely moving subject (Fanon, 84). This epidermal schema is one of the key psychological concepts in play when we consider how a racialized psyche differs from a white one.

Epidermalization is the name given to the process by which a racialized individual internalizes externally ascribed stereotypes due to projection onto the individual's body (Fanon, 84). One of the impacts of this process is the existence of a racialized individual "in whose soul an inferiority complex has been created by the death and burial of its local cultural originality" (Fanon, 9). It is not difficult to extrapolate the immensely destructive impacts such a process can have on the individuality of a racialized individual.

Fanon criticizes the model of psychoanalysis that characterized his time and argues against the capacity of European psychological theory to understand the colonial situation. This model, heavily influenced by ego psychology, focused primarily on individual internal conflicts and development, often neglecting the impact of sociopolitical factors like colonialism. He claims, using the example of the Malagasy, that colonial rule destroys the psychological mechanisms that a society has before it is touched by oppression and creates further adverse effects, such as the aforementioned inferiority complex that is often unconscious but still consequential (Fanon, 72). This changes the psychiatric treatment one must offer a Black patient, because a structural change like this requires not only psychoanalytic care individually, but also a group effort to change the social structures creating these complexes (Fanon, 74). Fanon acknowledges that the major psychological theories (which we now consider the foundations and facets of modern psychology) were not conceptualized with a racialized individual in mind at all; this spells trouble when we consider the psychic differences highlighted above (Fanon, 117). He places the cultural situation the colonized are in at the centre of every manifestation of neuroses wherein "a constellation of postulates, a series of propositions that slowly and subtly...work their way into one's mind and shape one's view of the world of the group to which one belongs" (Fanon, 152).

Fanon, while working as a psychiatrist, attempted to provide a psychiatric setting where people who would have been cut off from society and ultimately from themselves might be healed via alternative social interactions and activities, dubbed sociotherapy. Instead of making the patients subject to the power of a

medical system that they patronized, the intention behind these practices was to promote self-determination and community-building among the patients—a ‘reconstitution’ of the social order. When he worked in Algeria, he realized that the same assessments and therapies would not work for Muslim patients (as compared to European women) and learnt about their cultural emphasis on religion, family and storytelling as a form of entertainment (Novey; Sikuade). Accordingly, Fanon and his colleagues altered and oriented their sociotherapy to these interests (Novey; Sikuade).

Reconciling Fanon’s Framework and Contemporary Mental Healthcare

Fanon’s work takes place in a timeframe and socio-political context that vastly differs from the current Western and global state of affairs. It is valid to ask, ‘How does this fit into the picture of the current problems?’. The problem of racism is not nascent and cannot be said to be a thing of the past. Globalization has brought more people and cultures in contact than ever before, which can undoubtedly be construed as a sign of our immense progress. However, it also constitutes a site for newer, perhaps more covert, forms of racism (Hale; Ahlberg et al.) as well as an opportunity for the countries of the Global North to engage in neo-colonial tactics, resulting in greater inequalities between them and the Global South (Lodigiani). Additionally, the traumatic histories of racism and colonial violence are embedded within the cultural identities of the communities impacted, and to say that decisions made for and by individuals from these communities are not informed by this past, directly or indirectly, would be unintuitive and untrue. Even the biomedical model acknowledges the existence of intergenerational trauma, wherein trauma experienced by the older generation is “translated into a genetic adaptation that can be passed on to successive generations” (American Psychological Association). So, while our context is not identical to Fanon’s, it cannot be said that his theories surrounding racial and colonial identities are irrelevant, because discriminatory histories and practices have had an immense, and sometimes even irreversible, impact on generations past and present. The disproportionate impacts outlined in the first section highlight some of the disparities that still exist in regard to mental healthcare.

It is evident how Fanon’s focus on neuroses in a “cultural constellation” (Wynter) addresses the contemporary need for cultural competence in mental healthcare settings. His consideration of socio-political factors bolsters the psy-

chological liberation of oppressed groups and individuals as it affirms the distress brought about by harmful social factors, unlike the biomedical approach that puts the locus of control and deviant behaviours in the individual himself. Additionally, Fanon's acknowledgement that many foundational psychological theories are based on white people may be paralleled by the high rates of skewed samples in psychological studies even today, with most samples being drawn from the Western, Educated, Industrialized, Rich and Democratic (or "WEIRD") demographic (Henrich).

Conclusion

The modern biomedical model frames the ongoing mental health crisis in largely individual physiological terms and fails to adequately consider the impact of race, culture, colonization and oppression on the everyday life and emotions of large communities of individuals, while increasing essentialist thinking and stigma associated with having a mental disorder. Fanon, as a psychiatrist, was uniquely placed in a position of care wherein he was oppressed himself; this propelled him to consider the sociological forces that underlie psychological disorders. A careful consideration of these forces will address a large number of criticisms to the current mental healthcare framework. Frantz Fanon's work serves as a powerful reminder that mental health cannot be divorced from the social, cultural, and political contexts in which it exists. It is incumbent upon mental health professionals, policymakers, and society as a whole to embrace Fanon's call for a more humanistic and culturally competent approach to mental healthcare. Only then can we hope to create a system that truly addresses the needs of all individuals, regardless of their race, culture, or background.

Works Cited

- Ahlberg, B. M., et al. "Invisibility of Racism in the Global Neoliberal Era: Implications for Researching Racism in Healthcare." *Frontiers in Sociology*, vol. 4, 14 Aug. 2019. *Frontiers*, <https://doi.org/10.3389/fsoc.2019.00061>
- American Psychological Association. "Facing the Divide: Psychology's Conversation on Race and Health." www.apa.org/education-career/undergrad/diversity/facing-divide-instructor-guide.pdf
- Bredström, Anna. "Culture and Context in Mental Health Diagnosing: Scrutinizing the DSM-5 Revision." *The Journal of Medical Humanities*, vol. 39, no. 1, 2018, pp. 29–44. <https://doi.org/10.1007/s10912-017-9501-1>
- Comas-Díaz, Lillian, et al. "Racial Trauma: Theory, Research, and Healing: Toward Healing from Racial Oppression." *American Psychologist*, vol. 74, no. 1, 2019, pp. 85–100.
- Dyck, Arthur J. "Overdiagnosing, Overtreating, and Overmedicalizing Behavior and Feelings." *Library of Public Policy and Public Administration*, Springer, 2019, pp. 83–103. https://doi.org/10.1007/978-3-030-21707-5_5
- Fanon, Frantz. *Black Skin, White Masks*. Pluto Press, 2008.
- Franklin, Anderson J., et al. "Racism and Invisibility." *Journal of Emotional Abuse*, vol. 6, no. 2–3, Oct. 2006, pp. 9–30. https://doi.org/10.1300/j135v06n02_02
- Hale, Lindsay. "Globalization: Cultural Transmission of Racism." *Race, Gender & Class*, vol. 21, no. 1/2, 2014, pp. 112–25. <http://www.jstor.org/stable/43496963>
- Henrich, Joseph. *The WEIRD People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux, 2020.
- Herrera-Peco, I., et al. "The Role of Social Media as a Resource for Mental Health Care." *European Journal of Investigation in Health, Psychology and Education*, vol. 13, no. 6, 2023, pp. 1026–28. <https://doi.org/10.3390/eji13060078>
- Jacob, K. S., et al. "Diagnostic and Statistical Manual-5: Position Paper of the Indian Psychiatric Society." *Indian Journal of Psychiatry*, vol. 55, no. 1, 2013, pp. 12–30. <https://doi.org/10.4103/0019-5545.105500>

- Kirmayer, Laurence J. "Culture, Context and Experience in Psychiatric Diagnosis." *Psychopathology*, vol. 38, no. 5, 2005, pp. 191-196.
- Kirmayer, Laurence J. "Beyond the 'New Cross-Cultural Psychiatry': The Cultural Neuroscience of Mental Illness." *Transcultural Psychiatry*, vol. 46, no. 4, 2009, pp. 576-93.
- Kriegler, Susan. "A Social Constructivist Perspective on the Potential Relevance of Selected DSM-5 Disorders for South African Children and Youth." *Children & Society*, vol. 29, no. 6, 2015, pp. 604-14. <https://doi.org/10.1111/chso.12097>
- Kvaale, E. P., et al. "The 'Side Effects' of Medicalization: A Meta-Analytic Review of How Biogenetic Explanations Affect Stigma." *Clinical Psychology Review*, vol. 33, no. 6, 2013, pp. 782-94. <https://doi.org/10.1016/j.cpr.2013.06.002>
- Lodigiani, Irene. "From Colonialism to Globalisation: How History Has Shaped Unequal Power Relations Between Post-Colonial Countries." *Glocalism: Journal of Culture, Politics and Innovation*, 2020. <https://glocalismjournal.org/from-colonialism-to-globalisation-how-history-has-shaped-unequal-power-relations-between-post-colonial-countries/>
- Mackay, R., and J. R. Feagin. "'Merciless Indian Savages': Deconstructing Anti-Indigenous Framing." *Sociology of Race and Ethnicity*, vol. 8, no. 3, 2022, pp. 367-81. <https://doi.org/10.1177/23326492221112040>
- Mandalaywala, T. M., et al. "Essentialism Promotes Racial Prejudice by Increasing Endorsement of Social Hierarchies." *Social Psychological and Personality Science*, vol. 9, no. 4, 2018, pp. 461-69. <https://doi.org/10.1177/1948550617707020>
- Marsella, Anthony J. "Ethnocultural Aspects of PTSD: Issues, Research, and Clinical Applications." *Traumatology*, vol. 18, no. 4, 2012, pp. 66-77.
- McGuire, T. G., and J. Miranda. "New Evidence Regarding Racial and Ethnic Disparities in Mental Health: Policy Implications." *Health Affairs*, vol. 27, no. 2, 2008, pp. 393-403. <https://doi.org/10.1377/hlthaff.27.2.393>
- "Mental Disorders." World Health Organization, 8 June 2022, www.who.int/news-room/fact-sheets/detail/mental-disorders
- Mills, China. *Decolonizing Global Mental Health: The Psychiatrization of the Majority World*. Concepts for Critical Psychology, 2014.
- "News, August 2011. The Society's response to DSM-5; working memory

- ‘side-effects’; challenging public scepticism of science; and more” The Psychologist, British Psychological Society, Aug. 011, www.bps.org.uk/psychologist/news-august-2011
- Novey, Will. “Revisiting Frantz Fanon: His Life and Legacy on Race, Colonization, and Psychiatry.” *American Journal of Psychiatry Residents’ Journal*, vol. 16, no. 4, 2021, pp. 11–12.
- Okasha, Ahmed. “Globalization and Mental Health: A WPA Perspective.” *World Psychiatry*, vol. 4, no. 1, 2005, pp. 1–2.
- Panchal, Nirmita, et al. “The Implications of COVID-19 for Mental Health and Substance Use | KFF.” KFF, 25 Apr. 2023, www.kff.org/mental-health/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use.
- Rahman, M. M. “Questioning the Binaries between the Civilized and Uncivilized: A Feminist Reading of Hardy’s Tess and Thiong’o’s Wanja.” *Common Ground Research Networks*, 2023. <https://doi.org/10.18848/2327-7912/CGP/v21i02/21-37>
- Sartre, Jean-Paul. *The Philosophy of Jean-Paul Sartre*. United Kingdom, Knopf Doubleday Publishing Group, 2003.
- Sikuade, Adedapo. “Fifty Years after Frantz Fanon: Beyond Diversity.” *Advances in Psychiatric Treatment*, vol. 18, no. 1, 2012, pp. 25–31.
- Strakowski, S. M., et al. “Racial Influence on Diagnosis in Psychotic Mania.” *Journal of Affective Disorders*, vol. 39, no. 2, 1996, pp. 157–62. doi: 10.1016/0165-0327(96)00028-6
- Sue, Derald Wing. *Microaggressions and Marginality: Manifestation, Dynamics, and Impact*. John Wiley & Sons, Incorporated, 2010. <https://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=588930>
- Tribe, Rachel, and Sue Melliush. “Globalization, Culture and Mental Health.” *International Review of Psychiatry*, vol. 26, no. 5, 2014, pp. 535–37.
- Vernor, Dale. “PTSD Is More Likely in Women Than Men.” *National Alliance on Mental Illness*, 4 Sept. 2024, www.nami.org/stigma/ptsd-is-more-likely-in-women-than-men
- Vos, Johannes, et al. “Existential Therapies: A Meta-Analysis of Their Effects on Psychological Outcomes.” *Journal of Consulting and Clinical Psychology*, vol. 83, no. 1, 2015, pp. 115–28.
- Walker, E. R., et al. “Mortality in Mental Disorders and Global Disease Burden

- Implications.” JAMA Psychiatry, vol. 72, no. 4, 2015, pp. 334–41.
- Williams, David R., and Selcuk A. Mohammed. “Discrimination and Health: A Systematic Review.” Medical Care Research and Review, vol. 66, no. 6_suppl1, 2009, pp. 183S-237S.
- Wynter, Sylvia. “Towards the Sociogenic Principle.” Edited by Mercedes Durán-Cogan and Antonio Gómez-Moriana, University of Minnesota Press, 1999. trueleappress.wordpress.com/wp-content/uploads/2017/10/wynter-towards-the-sociogenic-principle.pdf

